



HAL
open science

CONTRIBUTIONS A LA MAITRISE STATISTIQUE DES PROCESSUS INDUSTRIELS MULTIVARIES

Teodor Tiplica

► **To cite this version:**

Teodor Tiplica. CONTRIBUTIONS A LA MAITRISE STATISTIQUE DES PROCESSUS INDUSTRIELS MULTIVARIES. Génie des procédés. Université d'Angers, 2002. Français. NNT: . tel-00739919

HAL Id: tel-00739919

<https://theses.hal.science/tel-00739919>

Submitted on 9 Oct 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**CONTRIBUTIONS A LA
MAITRISE STATISTIQUE DES PROCESSUS INDUSTRIELS
MULTIVARIÉS**

THESE DE DOCTORAT

Spécialité: Sciences et Technologies Industrielles

ECOLE DOCTORALE D'ANGERS

Présentée et soutenue publiquement

Le 03 Octobre 2002

A l'Institut des Sciences et Techniques de l'Ingénieur d'Angers

Par Teodor TIPLICA

Devant le jury ci-dessous:

René SOENEN	Examineur	Professeur à l'Université Claude Bernard Lyon 1
Zohra CHERFI	Rapporteur	Maître de Conférences HDR à l'Université de Technologie de Compiègne
Jean-François AUBRY	Rapporteur	Professeur à l'Institut National Polytechnique de Lorraine
Alain BARREAU	Examineur	Professeur à l'Université d'Angers
Abdessamad KOBI	Examineur	Maître de Conférences à l'Université d'Angers
Ioan BACIVAROV	Examineur	Professeur à l'Université Polytechnique de Bucarest

Directeur de thèse: **Alain BARREAU**
Co-encadrant: **Abdessamad KOBI**

Laboratoire: **Laboratoire en Sûreté de Fonctionnement, Qualité et Organisation**
62, avenue Notre Dame du Lac
49000 ANGERS

REMERCIEMENTS

Le travail présenté s'inscrit dans le cadre des activités de recherche menées au laboratoire LASQUO (Laboratoire en Sécurité de fonctionnement, Qualité et Organisation) à l'ISTIA (l'Institut des Sciences et Techniques de l'Ingénieur d'Angers), Université d'Angers.

Je tiens tout d'abord à remercier particulièrement Monsieur Abdessamad KOBI, Maître de Conférence à l'Université d'Angers, qui a dirigé de près l'ensemble de mes travaux, pour sa constante disponibilité, pour ses conseils, son soutien et ses nombreuses relectures de ce mémoire.

Je remercie Monsieur Alain BARREAU, Président de l'Université d'Angers, pour avoir accepté de diriger cette thèse malgré les nombreuses activités dont il a la charge ainsi que pour m'avoir suivi et guidé tout au long de ce travail.

Je remercie Monsieur René SOENEN, Professeur à l'Université Claude Bernard Lyon 1, pour l'honneur qu'il me fait de présider le jury.

Je tiens à remercier Monsieur Jean-François AUBRY, Professeur à l'Institut National Polytechnique de Lorraine, pour m'avoir fait l'honneur d'être rapporteur de cette thèse et pour ses remarques constructives.

J'adresse mes remerciements à Madame Zohra CHERFI, Maître de Conférence HDR à l'Université de Technologie de Compiègne, qui a accepté d'être rapporteur de cette thèse et pour ses remarques constructives.

Je remercie Monsieur Ioan BACIVAROV, Professeur à l'Université "POLITEHNICA" de Bucarest, pour m'avoir encouragé de poursuivre une carrière dans l'enseignement, pour l'opportunité qu'il m'a donné de venir en France dans le cadre du Programme TEMPUS dont il a été le coordonnateur et pour l'honneur qu'il me fait en participant à cette commission d'examen.

Je remercie Monsieur Jacques RENARD, Enseignant à l'Université d'Angers, qui a contribué par sa lecture attentive à améliorer ce document.

J'adresse mes remerciements et ma profonde reconnaissance à tous les membres du laboratoire avec qui j'ai pu collaborer durant ces travaux, et notamment à Fabrice GUERIN, Bernard DUMON, Mihaela BARREAU, Alexis TODOSKOFF, Ridha HAMBALI et Christian ROBLEDO pour leurs conseils et leur soutien.

Enfin, je remercie ma fiancée Anita pour m'avoir aidé et soutenu dans les moments les plus difficiles et toute ma famille pour leur soutien et leurs encouragements permanents.

SOMMAIRE

CHAPITRE 1

1.1 Introduction	1
1.2 La "Qualité" nécessité économique et objectif prioritaire des entreprises	2
1.3 Qu'est-ce que le Contrôle Statistique de la Qualité ?	3
1.4 La Maîtrise Statistique de Processus	4
1.5 - Les cartes de contrôle	6
1.5.1 Les cartes de contrôle univariées	6
1.5.2 Les cartes de contrôle multivariées	8
1.6 Conclusion	19

CHAPITRE 2

2.1 Introduction	21
2.2 Panorama des principes de détection	22
2.3 Description et analyse des méthodes existantes	23
2.3.1 Méthode de régression des variables	23
2.3.2 Méthode basée sur l'utilisation de la statistique t de Student	25
2.3.3 Décomposition MYT de la carte de contrôle T^2 de Hotelling	27
2.3.4 Algorithme BSA (Backward Selection Algorithm) et méthode de l'hyperplan	31
2.3.5 Méthode de calcul des contributions des variables	35
2.3.6 Carte de contrôle sélection de la cause	38
2.3.7 Carte de contrôle multivariée des profils	40
2.3.8 Les méthodes de projection	42
2.3.9 Méthodes de détection directionnelles	45
2.4 Conclusion	49

CHAPITRE 3

3.1 Introduction	51
3.2 Analyse discriminante	52
3.2.1 Formulation du problème et notations	52
3.2.2 Axes factoriels discriminants. Métrique discriminante	55
3.2.3 Règles d'affectation ou de classement	56
3.2.4 Qualité des règles de classement	58
3.3 Filtrage numérique	59
3.3.1 Introduction	59
3.3.2 Types de filtres numériques	60
3.4 Méthode FNAD (Filtrage Numérique et Analyse Discriminante)	62
3.4.1 Introduction	62
3.4.2 Types de causes assignables étudiées	66
3.4.3 Analyse spectrale des causes assignables	68
3.4.4 Module d'analyse "off-line"	74
3.4.5 Module bibliothèque de classes	102
3.4.6 Module d'analyse "on-line"	104
3.5 Application industrielle	117
3.5.1 Introduction	117
3.5.2 Le processus de fabrication du polyéthylène	117
3.5.3 Modélisation mathématique du processus de fabrication	119
3.5.4 Choix des variables suivies	120
3.5.5 Création de la base de données	121
3.5.6 Module d'analyse "off-line" (Analyse exploratoire de données)	123
3.5.7 Bibliothèque des classes	126
3.5.8 Module d'analyse "on-line"	127
3.6 Conclusion	131

CHAPITRE 4

4.1 Introduction	133
4.2 Carte de Contrôle par Filtrage Numérique (CCFN)	134
4.2.1 Introduction	134
4.2.2 Principe de construction de la carte CCFN	135
4.2.3 Carte multidimensionnelle de contrôle par filtrage numérique CMCFN	140
4.2.4 Comparaison de la carte de contrôle EWMA avec la CCFN	142
4.2.5 Evaluation de la performance de la carte de contrôle CCFN	145
4.2.6 Exemples d'application de la CCFN	147
4.3 Carte de contrôle spectrale	150
4.3.1 Introduction	150
4.3.2 Principe de construction de la carte de contrôle spectrale	151
4.3.3 Etude du spectre des causes assignables	153
4.3.4 Taille de la fenêtre temporelle	156
4.3.5 Détection des causes assignables avec la carte de contrôle spectrale	158
4.3.6 Analyse de l'efficacité de la carte de contrôle spectrale	160
4.4 Conclusion	161
5. Conclusion générale et perspectives	163
Annexes	165
A.1 Evénements majeurs dans le domaine de la Qualité	165
A.2 Calcul des coefficients de régression biaisés	170
A.3 Carte de contrôle pour des observations couplées	171
A.4 Calcul des axes factoriels discriminants	175
A.5 Formule récurrente y en fonction de x pour un filtre ARMA	175
A.6 Calcul de la POM1 par la méthode des chaînes de Markov	176
Références	179

Chapitre 1

- 1.1 Introduction
- 1.2 La "Qualité" nécessité économique et objectif prioritaire des entreprises
- 1.3 Qu'est-ce que le Contrôle Statistique de la Qualité ?
- 1.4 La Maîtrise Statistique de Processus
- 1.5 Les cartes de contrôle
- 1.6 Conclusion

1.1 Introduction

Dans ce premier chapitre nous essayerons d'abord de donner une vision d'ensemble sur la "Qualité" et de tracer l'évolution des principaux concepts philosophiques la définissant. Les différentes méthodes d'amélioration de la qualité seront passées en revue et un intérêt particulier sera accordé au Contrôle Statistique de la Qualité (CSQ). Ensuite, nous focaliserons notre attention sur des aspects relevant de la Maîtrise Statistique de Processus (MSP) et principalement sur les cartes de contrôle. Nous nous contenterons de faire une présentation sommaire des principales cartes de contrôle univariées et nous soulignerons leurs contributions à l'amélioration de la qualité. En final, les perspectives ouvertes dans le domaine de la MSP par le progrès de la technologie électronique et informatique et l'évolution vers un contrôle statistique multivarié seront analysées. Les cartes de contrôle multivariées les plus connues et leurs principes de construction seront présentés.

1.2 La "Qualité" nécessité économique et objectif prioritaire des entreprises

Dans le contexte économique actuel, les entreprises sont confrontées à une terrible épreuve : gagner des clients et surpasser leurs concurrents. Afin d'atteindre ces objectifs, les entreprises doivent faire des efforts pour assurer et satisfaire les besoins de leurs clients.

Notre objectif, en tant qu'entreprise [Wal-Mart], est d'avoir une stratégie orientée client qui n'est pas seulement la meilleure, mais légendaire. (Sam WALTON)

Si chacun d'entre nous, peut continuer à effectuer un travail dans cette entreprise [Chrysler] c'est grâce à la qualité, à la productivité et à la satisfaction du client. (Lee IACocca)

Dans ce cadre, la qualité des produits et des services est devenue parmi les facteurs de décision, un des plus importants dans le choix des consommateurs. Par conséquent, comprendre et améliorer la qualité est un point clé pour le succès dans les affaires, pour la croissance économique et la compétitivité d'une entreprise. L'introduction des méthodes formelles pour le contrôle et l'amélioration de la qualité a été ressentie comme une nécessité absolue.

L'évolution chronologique d'événements majeurs dans le domaine de la qualité est présentée dans le tableau 1 (voir l'annexe A.1).

Mais, qu'est-ce que la "Qualité" ? Il est difficile de donner une réponse simple à cette question. David Garvin [37] considère que la qualité présente huit composantes (*La performance, La fiabilité, La durabilité, La réparabilité, L'esthétique, Les fonctionnalités, L'image de marque, La conformité aux normes*). La définition classique de la qualité tient compte du fait que les produits et les services doivent satisfaire les exigences de ceux qui les utilisent :

"Qualité signifie adéquation à l'utilisation."

Dans leur démarche vers la **Qualité Totale**, les entreprises doivent se définir une solide *Stratégie Qualité*. La volonté et la détermination de la direction jouent un rôle important pour le succès et pour l'impact de la stratégie qualité dans une entreprise. Le Management de la Qualité Totale (*MQT*) est un concept de structure de gestion qui crée un cadre favorable à l'application et l'utilisation des méthodes statistiques. La qualité se construit autour d'un *Système Qualité* à l'aide des outils comme la MSP (*Maîtrise Statistique de Processus*), l'AMDEC (*Analyse des Modes de Défaillance, de leurs Effets et de leur Criticité*), les Plans d'Expériences, la **QFD** (*Quality Function Deployment*), etc.

Actuellement, on parle de plus en plus de la méthode **Six Sigma** qui a ses origines chez Motorola où elle fut lancée en 1987. Six Sigma est une méthodologie fondée sur trois principes : réduire le nombre de rebuts, augmenter la satisfaction des clients, et améliorer les processus. Le programme Six Sigma apporte des résultats quantifiables et mesurables financièrement. Selon la méthodologie Six Sigma la réduction de la variabilité est un élément clé pour l'amélioration des processus. La MSP,

par l'utilisation de l'outil statistique, fournit une contribution très importante à cette évolution centrée sur la production autour d'une valeur cible et en réduisant, en même temps, la dispersion autour de cette valeur cible. Ainsi, le raisonnement fondé sur des pourcentages de pièces non conformes a été remplacé par la notion de pièces par million non-conformes (*ppm*).

Ayant en vue cet aspect, il est préférable de donner une définition plus moderne de la qualité :

"La Qualité est inversement proportionnelle à la variabilité."

Selon cette définition, la qualité d'un produit sera d'autant plus grande que la variabilité de ses caractéristiques sera faible.

1.3 Qu'est-ce que le Contrôle Statistique de la Qualité ?

Le CSQ (Contrôle Statistique de la Qualité) recouvre plusieurs aspects de la statistique industrielle comme :

- ✓ le contrôle par échantillonnage
- ✓ la maîtrise statistique de processus
- ✓ les plans d'expériences
- ✓ l'analyse de capacité

Le contrôle par échantillonnage est utilisé pour avancer des décisions concernant l'acceptation ou non d'un lot de produits. Il est souvent utilisé en contrôle de fin de fabrication ou encore en contrôle de réception. De façon générale, ce contrôle s'effectue à l'aide d'un plan d'échantillonnage. Les plans d'échantillonnage les plus connus sont : le plan d'échantillonnage simple; le plan d'échantillonnage double; le plan d'échantillonnage multiple et le plan d'échantillonnage progressif.

A partir des années 1980 le contrôle par échantillonnage est moins utilisé. Deming [20] considère que le contrôle par échantillonnage "*est effectué trop tard, est trop cher et finalement non-efficace*". Une des raisons pour laquelle cette technique est moins utilisée à présent est que les méthodes statistiques sont plus efficaces quand elles sont appliquées dans la phase de démarrage du processus de production y compris la conception et le développement. Cela ne veut pas dire que le contrôle par échantillonnage est complètement sans objet.

La MSP est utilisée pour visualiser l'évolution temporelle d'un processus de production et pour détecter les changements susceptibles de modifier ses performances. Une présentation plus élaborée de cet aspect du CSQ sera faite dans le paragraphe 1.4.

Les plans d'expériences sont employés afin de déterminer quels sont les facteurs (variables) influant sur la qualité d'un processus ou d'un produit. Le principe qui est à la base de la théorie des plans d'expérience pourrait se résumer ainsi : un changement systématique des facteurs contrôlables afin de déterminer leurs effets sur les paramètres du produit. Ce principe a été appliqué dans de nombreux secteurs comme l'électronique, l'industrie des semi-conducteurs, l'aérospatiale, l'automobile, l'industrie alimentaire, l'industrie pharmaceutique, l'industrie chimique, etc. Les plans d'expériences sont des outils importants de contrôle qualité "off-line" puisqu'ils sont souvent utilisés au moment de développement ou de démarrage d'une activité de production contrairement aux méthodes dites "on-line" c'est-à-dire en cours de fabrication. Ils jouent un rôle

crucial dans la réduction de la variabilité. Les plans d'expériences et les méthodes de maîtrise statistique de processus sont étroitement liés. La MSP est une méthodologie qu'on peut définir comme étant "*passive*" car le processus est surveillé dans l'attente d'une certaine information indiquant la nécessité d'une intervention. Si le processus est sous contrôle, les observations ne fournissent pas beaucoup d'informations. Par contre, les plans d'expériences constituent une méthode statistique "*active*". On effectue une série de tests sur le processus en changeant les entrées et en observant ce qui se passe à la sortie. Cela, donne une grande quantité d'informations qui pourra être utilisée pour l'amélioration du processus.

Genichi Taguchi [116-121] a remarquablement développé et vulgarisé les plans d'expériences. La méthode de Taguchi a été utilisée au début des années 1980 aux Etats-Unis. Taguchi propose une méthodologie qui rend les produits et les processus plus robustes aux variations générées par des facteurs difficilement contrôlables.

L'objectif de l'analyse de capabilité est de déterminer si le processus est capable de respecter les spécifications prédéfinies pour une certaine caractéristique de qualité donnée. L'estimation de la capabilité d'un processus peut se faire en terme de distribution de probabilité. Dans cette optique l'étude de capabilité pourra se faire sans avoir besoin des spécifications pour la caractéristique de qualité analysée. Sinon, on peut exprimer la capabilité d'un processus comme le pourcentage des dépassements de la limite de spécification. L'analyse de la capabilité d'un processus représente une étape vitale dans tout programme qui vise l'amélioration de la qualité. Les résultats d'une analyse de capabilité peuvent être utilisés ultérieurement pour :

- ✓ prédire dans quelle mesure le processus va respecter les limites de spécification imposées.
- ✓ assister les concepteurs des produits dans la sélection ou la modification d'un processus.
- ✓ aider à établir un intervalle d'échantillonnage pour le suivi à l'aide des cartes de contrôle.
- ✓ fixer les contraintes de performance pour un nouvel équipement.
- ✓ choisir parmi plusieurs fournisseurs.
- ✓ planifier l'ordre des étapes et sous-étapes d'un processus quand cela peut avoir un effet sur le respect de la limite de spécification imposée.
- ✓ réduire la variabilité dans un processus de production.

Ainsi, l'analyse de la capabilité est une technique qui trouve plusieurs applications dans tout le cycle de fabrication d'un produit, à partir de la conception, le choix du fournisseur, la planification de la production et jusqu'à la réalisation effective du produit.

1.4 La Maîtrise Statistique de Processus

Sous le nom de MSP (Maîtrise Statistique de Processus) on regroupe toute méthode statistique conçue pour détecter des changements dans un processus. Selon Pearson [91, 92], la modélisation mathématique et l'analyse de la performance des cartes de contrôle ont commencé en 1932 après une visite de Shewhart [108] en Angleterre. Au début, ces méthodes ont été utilisées parce qu'elles étaient simples, faciles à mettre en œuvre et donnaient des bons résultats. Par exemple, l'étendue pour estimer la dispersion d'un processus est une pratique courante même aujourd'hui. Il ne faut pas oublier surtout, qu'à l'époque, les échantillons étaient

mesurés à la main à des intervalles de temps relativement éloignés vue la difficulté de l'échantillonnage et les instruments de mesure peu performants.

Jusqu'à la fin des années 1970, beaucoup d'améliorations ont été proposées mais peu de chercheurs s'intéressaient vraiment à ces méthodes statistiques (Woodall et Montgomery [131]). La MSP reposait encore à cette époque sur des méthodes univariées. Le problème est que la qualité d'un produit ou d'un processus est, comme nous l'avons précisé dans un paragraphe antérieur, une propriété ayant plusieurs aspects et elle doit être traitée par conséquent ainsi. L'activité de recherche a connu une dynamique très importante à partir des années 1980. L'augmentation de l'intérêt pour ces méthodes a été provoquée, sans aucun doute, par la révolution dans le domaine de la qualité suite à l'ouverture du marché international et à la concurrence entre les entreprises. La notion de "Maîtrise de Processus" n'a pas toujours de signification claire pour les ingénieurs et cela peut causer des difficultés de communications. Les ingénieurs ayant une formation en électronique, automatique, mécanique ou chimie associent très souvent cette notion avec le contrôle prédictif, le contrôle adaptatif ou le contrôle proportionnel-intégral-différentiel décrit dans des livres de théorie des systèmes linéaires. Par contre ceux qui ont une formation en statistique ou ingénierie industrielle associent cette notion avec les cartes de contrôle et le suivi du processus. Des méthodes statistiques de contrôle pour les ingénieurs ont été diffusées dans toutes les domaines industriels par des spécialistes comme Fisher, Youden, Box et Tukey. Ces "statisticiens" ont compris la façon dont les spécialistes et les ingénieurs avaient l'habitude de voir les choses et ils ont développé des méthodes statistiques pour traiter leurs problèmes concrets au lieu d'élaborer une "belle" théorie académique et essayer ensuite de trouver une application sur mesure.

Le milieu industriel d'aujourd'hui est beaucoup plus favorable à l'application de la MSP. Les mesures "on-line", l'archivage des données, l'analyse assistée par l'ordinateur sont souvent employés. Cela a changé totalement la nature des données disponibles pour le contrôle du processus. Dans certains cas, les données sont disponibles à chaque seconde pour des centaines de variables de processus. Cela pose aussi beaucoup de problèmes et nous rappelons ici seulement quelques uns parmi les plus importants :

- ✓ *La dimensionnalité* : dans beaucoup d'applications pratiques en temps réel, une très grande quantité de données est enregistrée sur un nombre élevé des variables. Les opérateurs sont donc obligés de visualiser seulement un faible nombre de variables clés sur l'écran d'ordinateur.
- ✓ *La colinéarité* : même si plusieurs centaines de variables sont mesurées, toutes ne sont pas indépendantes. Ces variables sont liées à un nombre limité d'évènements pouvant influencer le système et par conséquent elles sont fortement corrélées. Néanmoins, chaque variable contrôlée est utile car elle apporte une petite quantité supplémentaire d'informations au sujet de chaque événement qui se produit dans le système.
- ✓ *Le bruit* : les variables mesurées sont souvent enregistrées avec des erreurs (erreurs de mesure ou erreurs d'échantillonnage). Par conséquent, l'utilisation directe des cartes de contrôle multivariées n'est pas recommandée. Tenant compte que dans chaque variable de processus se trouve une toute petite quantité d'informations concernant un certain événement se produisant dans le système, il est possible en utilisant des méthodes statistiques multivariées adéquates d'extraire cette information très précieuse.

- ✓ *Les données manquantes* : les grandes quantités d'enregistrements effectuées sur un processus peuvent présenter aussi des données manquantes (parfois jusqu'à 20% des données). Par conséquent, si on veut analyser un tel processus à l'aide des méthodes statistiques multidimensionnelles, on doit concevoir des algorithmes nous permettant d'opérer dans ces conditions.

Des cartes de contrôle classiques sont utilisées dans plusieurs applications en "temps réel". Très souvent la notion de "temps réel" n'est pas clairement définie. Elle peut représenter des prélèvements (ou des enregistrements) effectués à l'intervalle de quelques secondes ou de quelques heures. De multiples exemples peuvent être donnés dans une large palette de domaines industriels comme : l'industrie des semi-conducteurs, l'industrie alimentaire, l'industrie de l'automobile, les communications, l'analyse de l'image, la reconnaissance vocale et l'industrie chimique.

La gestion des données et des limites de contrôle en temps réel peut poser souvent des problèmes. La plus part des industriels s'accorde sur le fait que les applications d'envergure nécessitent des logiciels spécialisés effectuant des traitements statistiques et reliés à des systèmes d'acquisition en temps réel.

On peut constater aujourd'hui que plusieurs types de méthodes statistiques ont été intégrées dans les procédures de la MSP. On mentionne ici les méthodes de régression, l'analyse multivariée, les séries chronologiques, les méthodes non-paramétriques, pour donner seulement quelques exemples.

1.5 - Les cartes de contrôle

L'utilisation d'une carte de contrôle repose sur l'échantillonnage d'un processus qui évolue dans le temps. Il faut établir quelle sera la variable mesurée, la taille de l'échantillon, l'intervalle entre deux prélèvements et les règles de décisions à appliquer.

1.5.1 Les cartes de contrôle univariées

Le principe de base dans la construction de la plupart des cartes univariées repose sur la notion de test d'hypothèse. Il est généralement considéré que le modèle théorique du processus suit une certaine distribution de probabilité. Ensuite, l'hypothèse nulle (ou "sous-contrôle") pour ses paramètres (comme la moyenne ou la variance) est spécifiée. L'hypothèse alternative consistera alors dans le changement de la moyenne ou de la variance du processus.

La carte de contrôle Shewhart classique, a été conçue pour visualiser l'évolution de la moyenne d'un processus. Elle est constituée d'une ligne qui met en évidence la tendance centrale du processus et deux limites de contrôle (la limite supérieure de contrôle et la limite inférieure de contrôle). Les moyennes d'échantillons prélevés de la production y sont représentées en fonction du moment de prélèvement. La carte de contrôle met en évidence une situation "hors-contrôle" chaque fois qu'un point dépasse une des deux limites de contrôle. Les limites de contrôle sont choisies habituellement à $\pm 3\sigma_{\bar{x}}$ par rapport à la droite centrale, où $\sigma_{\bar{x}}$ est l'écart-type estimé de la moyenne d'échantillons.

Plusieurs méthodes ont été proposées pour améliorer la sensibilité de la carte de contrôle par rapport aux sauts de faible amplitude de la moyenne du processus. Il s'agit,

par exemple, de l'utilisation des règles d'alerte dans le cas d'une disposition particulière des points sur la carte, appelées aussi les règles de Western Electrics.

Modifier l'intervalle entre deux prélèvements est une autre solution à ce problème. La carte de contrôle basée sur l'intervalle d'échantillonnage variable (IEV) consiste à échantillonner le processus à une basse fréquence jusqu'au moment où il y a des indications sur la présence des causes assignables (par exemple des points situés à proximité de la limite de contrôle). Dans de telles situations la fréquence d'échantillonnage est augmentée. Ainsi, même si le nombre d'échantillons nécessaires pour détecter la situation hors-contrôle reste le même, le temps de détection est réduit. Si, après l'augmentation de la fréquence d'échantillonnage, il s'avère que le processus est toujours sous contrôle, on revient à la fréquence d'échantillonnage antérieure. L'avantage de la méthode (IEV) est de conserver la simplicité de l'inspection visuelle et de l'interprétation de la carte de contrôle Shewhart

Les cartes de contrôle CUSUM (*Cumulative SUM*) et EWMA (*Exponentially Weighted Moving Average*) ont été aussi conçues pour détecter de faibles changements dans un processus. Elles permettent l'accumulation de l'information dans le temps.

La carte CUSUM repose sur un modèle de processus dont les observations \mathbf{x}_k suivent une distribution normale de moyenne μ et d'écart-type σ . La carte consiste à représenter la somme cumulée de la différence entre les observations \mathbf{x}_k et la valeur moyenne μ , c'est-à-dire : $\sum_{k=1}^t (x_k - \mu)$, aux différents moments d'échantillonnage \mathbf{k} . Une dérive de la moyenne du processus est mise en évidence par une tendance ascendante ou descendante prononcée sur la carte de contrôle.

La carte de contrôle EWMA accumule les informations d'une manière différente par rapport à la carte CUSUM. On fixe la valeur de démarrage $\mathbf{y}_0 = \mu_0$ (où μ_0 est la moyenne du processus) et on affiche sur la carte la statistique :

$$y_k = r x_k + (1 - r) y_{k-1}, \quad 0 < r \leq 1$$

où \mathbf{k} représente l'instant d'échantillonnage et \mathbf{r} est un paramètre de pondération ayant des valeurs comprises entre 0 et 1. Un dérèglement est détecté à chaque fois que la statistique \mathbf{y}_k dépasse une limite de contrôle. La construction de la carte EWMA repose sur le choix du paramètre de pondération \mathbf{r} et de la POM (Période Opérationnelle Moyenne) qu'on souhaite avoir quand le processus est sous contrôle.

Il arrive souvent que les " caractéristiques qualité " ne soit pas mesurables (un contrôle visuel par exemple) ou encore qu'il est plus pratique et plus économique d'effectuer un contrôle qualitatif à l'aide de calibres pour classer les pièces conformes ou non aux spécifications. Ce type de contrôle porte le nom de contrôle par attributs et consiste à noter la présence d'un critère qualitatif quelconque (attribut) sur chaque unité contrôlée et de compter le nombre d'unités possédant ce critère sur l'ensemble d'unités contrôlées. Ce type de contrôle est plus rapide, plus facile et moins coûteux que le contrôle par mesures mais apporte toutefois une information plus réduite. Les principales cartes pour un contrôle par attributs sont : la carte P (carte de contrôle pour la proportion de non conformes); la carte NP (carte de contrôle pour le nombre d'unités non conformes); la carte C (carte de contrôle pour le nombre de non-conformités), la carte U (carte de contrôle pour le nombre de non-conformités par unité) et la carte D (ou carte du nombre total de démerites par unité contrôlée).

Il est très important, dans la MSP, de faire la distinction entre l'étape d'analyse rétrospective (*Etape I*) et l'étape de suivi du processus (*Etape II*). A l'Etape I, une certaine quantité de données est collectée. Pour une carte de contrôle pour les moyennes, par exemple, il est conseillé d'extraire au moins 20-25 échantillons de taille 4 ou 5. Ensuite, on vérifie si le processus est sous-contrôle et on estime les paramètres qui seront utilisés ultérieurement dans le calcul des limites de contrôle. A l'étape II, les performances statistiques sont quantifiées par la POM (Période Opérationnelle Moyenne) de la carte de contrôle. La POM représente le nombre moyen d'échantillons successifs prélevés jusqu'à ce que la situation hors-contrôle soit détectée sur la carte de contrôle.

1.5.2 Les cartes de contrôle multivariées

Les cartes de contrôle univariées ont eu une contribution remarquable à l'amélioration de la qualité des processus mais leur utilisation actuelle reste une réminiscence du passé. Ces cartes de contrôle, en faisant intervenir des concepts statistiques, ont contribué également au changement de la mentalité de ceux qui les ont utilisés à l'époque. Actuellement, suite à l'évolution remarquable de l'électronique et de l'informatique, on a le moyen de faire mieux car, il faut être conscient, que le processus de fabrication n'est jamais univarié. En effet, le processus de fabrication est régi par un grand nombre de variables ou des facteurs qu'il faut maîtriser simultanément afin de s'assurer le succès de la démarche mise en œuvre et l'obtention des bons résultats.

La maîtrise des processus industriels est, donc obtenue suite à l'observation permanente d'un grand nombre des facteurs (figure I.1). A l'entrée du processus, de nombreuses mesures sont effectuées sur les variables de processus et à la sortie, un grand nombre de caractéristiques de produit sont mesurées également. L'utilisation d'une carte de contrôle univariée de type Shewhart pour chaque variable de processus peut être une solution (variables indépendantes) mais dans le cas de variables corrélées, les résultats obtenus sont erronés car la probabilité associée à l'erreur de type I (l'erreur de première espèce) est plus grande. Les cartes de contrôle multivariées sont plus adéquates pour le suivi d'un processus multivarié. Leur principal avantage, par rapport aux cartes univariées, est la prise en compte de la corrélation existante entre les variables.

Hotelling [53], a été le premier à analyser un système de variables corrélées. Il a développé une procédure de contrôle basée sur le concept de "distance statistique", qui a été nommée la statistique T^2 de Hotelling. Dans le même temps, Mahalanobis a développé une distance statistique similaire qui a été nommée la distance de Mahalanobis. Les deux distances statistiques diffèrent par une constante. Nous allons décrire dans le prochain paragraphe la distance statistique T^2 de Hotelling car elle s'est avérée une très bonne métrique pour le contrôle statistique multivarié.

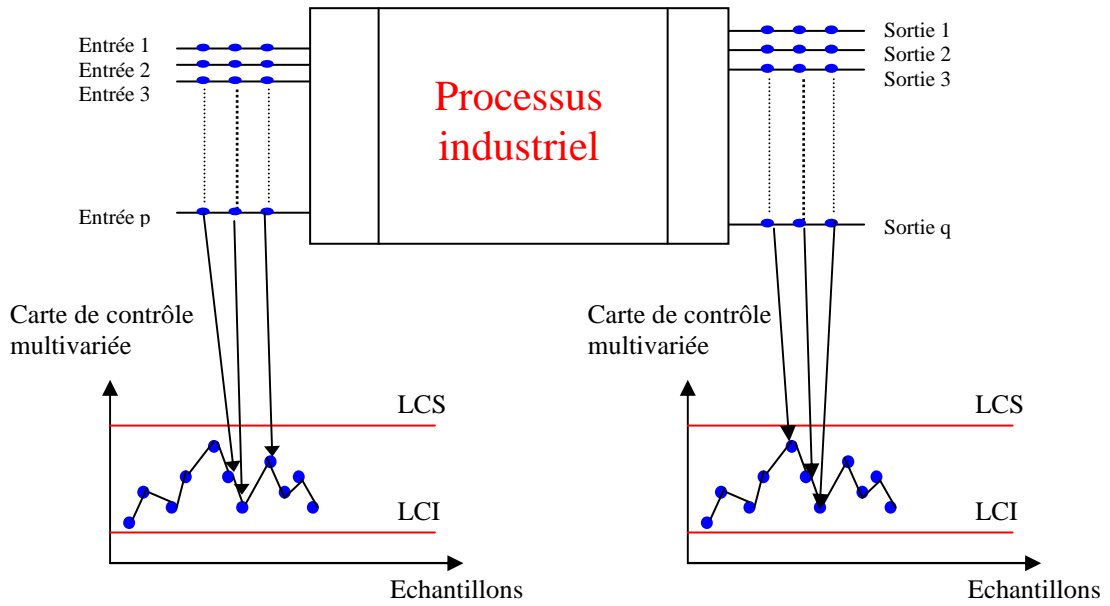


figure I.1 – Processus industriel multivarié

La distance statistique T^2 de Hotelling

Un ensemble d’observations effectuées sur deux variables indépendantes X_1 et X_2 peut être représenté comme un nuage de points dans un espace bidimensionnel. On suppose que l’on veut calculer la distance entre un point quelconque **A** de coordonnées (x_1, x_2) et le centre de gravité du nuage **M**. La distance entre les deux points se traduit toujours par un seul nombre, quelle que soit la dimension de l’espace. Le nombre d’unités métriques qui les sépare représente la distance Euclidienne entre les deux points. Mathématiquement, la distance entre le point **A** et le centre de gravité du nuage **M** (μ_1, μ_2) est donnée par l’équation suivante :

$$(x_1 - \mu_1)^2 + (x_2 - \mu_2)^2 = D^2 \tag{1.1}$$

Ainsi, on a transformé une observation binaire (x_1, x_2) en un seul nombre **D** qui représente la distance entre l’observation et le centre de gravité du nuage **M**. On remarque aussi que tous les points ayant une distance constante par rapport au centre de gravité du nuage se situent sur un cercle de centre **M** et de rayon égal à **D** (figure I.2.a).

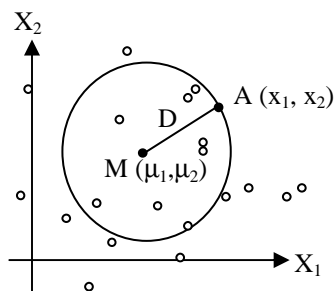


figure I.2.a - Distance Euclidienne

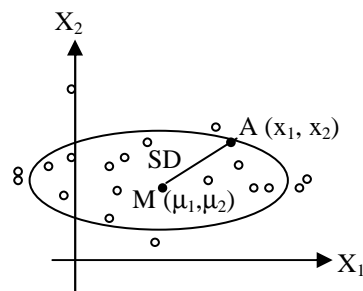


figure I.2.b - Distance Statistique

Cette métrique est insatisfaisante dans la plupart des cas parce qu'elle donne à chaque variable le même poids dans le calcul de la distance et parce qu'elle ne prend pas en compte les différences d'échelle entre les variables. Pour corriger ces inconvénients, les variables sont centrées et réduites en utilisant la transformation suivante :

$$\frac{(x_1 - \mu_1)}{\sigma_1} \quad \text{et} \quad \frac{(x_2 - \mu_2)}{\sigma_2}$$

En remplaçant ces valeurs dans l'équation (1.1) on obtient :

$$\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} = SD^2 \quad (1.2)$$

Cette métrique est connue sous le nom de "distance statistique". Tous les points satisfaisant l'équation (1.2) ont la même distance statistique par rapport au centre de gravité \mathbf{M} . La représentation graphique de l'équation (1.2) est une ellipse ayant les deux axes parallèles avec les axes de coordonnées (X_1 et X_2) comme dans la figure I.2.b.

En comparant les deux métriques, on constate que dans le cas de la distance statistique les axes de coordonnées n'ont pas d'unité de mesure. Cette propriété a des conséquences très importantes en pratique car on peut travailler avec des variables ayant des unités de mesure différentes. Si l'on regarde la figure I.2.b on peut constater que deux points peuvent avoir la même distance statistique, mais des distances Euclidiennes différentes. La différence majeure entre la distance statistique et la distance Euclidienne réside dans le fait que, dans le calcul de la distance statistique les variables ont des poids inversement-proportionnelles à leurs variances, alors que pour le calcul de la distance Euclidienne les deux variables ont les mêmes poids. Ainsi, une variable ayant une faible variabilité va contribuer d'une façon plus importante au calcul de la distance statistique qu'une variable ayant une grande variabilité. Si les deux variables ont la même variance il n'y a pas de différence entre la distance statistique et la distance Euclidienne.

Si les deux variables sont corrélées, l'axe principal et l'axe secondaire de l'ellipse ne sont plus parallèles avec les axes de coordonnées (X_1 , X_2). L'axe principal de l'ellipse sera la droite de régression d'une variable dans l'autre comme dans la figure ci-dessous :

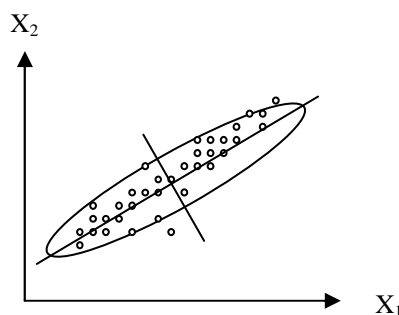


figure I.3 - Cas de deux variables corrélées

Selon la géométrie analytique, l'équation de l'ellipse dans ce cas est :

$$a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 = SD^2 \quad (1.3)$$

Dans cette équation le facteur $2a_{12}x_1x_2$ donne l'inclinaison de l'ellipse.

Une hypothèse fondamentale dans la construction de la statistique T^2 de Hotelling est que le vecteur d'observations suit une distribution normale multidimensionnelle. Si cette supposition est fondée, alors on peut écrire la distance entre le point $\mathbf{A}(x_1, x_2)$ et le centre de gravité du nuage à l'aide de la distribution normale bivariée. L'équation (1.3) devient après la standardisation des variables :

$$\frac{1}{1-\rho^2} \left[\frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right] = SD^2 \quad (1.4)$$

où ρ représente la corrélation entre les variables X_1 et X_2 .

En utilisant la notation matricielle, l'équation (1.4) peut être exprimée comme suit :

$$(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) = SD^2 \quad (1.5)$$

où $\mathbf{X}^T = (x_1, x_2)$ est le vecteur d'observations, $\boldsymbol{\mu}^T = (\mu_1, \mu_2)$ est le vecteur des moyennes et $\boldsymbol{\Sigma}^{-1}$ est l'inverse de la matrice de variance-covariance :

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

où: $\sigma_{12} = \sigma_{21}$ représente la covariance entre X_1 et X_2 .

L'équation (1.5) est connue comme la statistique T^2 de Hotelling :

$$T^2 = (\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) = SD^2 \quad (1.6)$$

L'équation de l'ellipse est obtenue par l'équation $T^2 = \mathbf{c}$, où \mathbf{c} est une constante. Tous les points à l'intérieur de l'ellipse ont une distance statistique inférieure à \mathbf{c} . La constante \mathbf{c} est considérée comme une limite de contrôle supérieure ou une valeur critique pour la région de contrôle multidimensionnelle. En fonction de la corrélation existante entre les variables, l'ellipse est plus ou moins penchée et plus ou moins allongée. Quand il n'y a pas de corrélation entre les variables, le suivi d'un processus à l'aide de la statistique T^2 est similaire au contrôle du processus utilisant des cartes Shewhart pour chaque variable.

La carte de contrôle T^2 de Hotelling

Si l'on connaît les paramètres du processus (vecteur des moyennes et la matrice variance-covariance) la statistique T^2 de Hotelling pour deux variables X_1 et X_2 est selon l'équation (1.4) :

$$\begin{aligned}
 T^2 &= \frac{1}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} \left[\sigma_2^2 (\bar{x}_1 - \mu_1)^2 + \sigma_1^2 (\bar{x}_2 - \mu_2)^2 - 2\sigma_{12} (\bar{x}_1 - \mu_1)(\bar{x}_2 - \mu_2) \right] = \\
 &= \frac{1}{1 - \rho^2} \left[\frac{(\bar{x}_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \left(\frac{\bar{x}_1 - \mu_1}{\sigma_1} \right) \left(\frac{\bar{x}_2 - \mu_2}{\sigma_2} \right) + \frac{(\bar{x}_2 - \mu_2)^2}{\sigma_2^2} \right]
 \end{aligned} \tag{1.7}$$

où \bar{x}_1 et \bar{x}_2 sont les moyennes d'échantillons de taille n prélevés par tirage aléatoire du processus, σ_1 , σ_2 et μ_1 , μ_2 sont respectivement les écarts-type et les moyennes de deux variables X_1 et X_2 ; σ_{12} et ρ sont respectivement la covariance et la corrélation de deux variables.

En fixant une valeur constante pour la partie gauche de l'équation (1.7) on obtient l'équation d'une ellipse. Montgomery [85] a démontré que, si l'on connaît les paramètres du processus, la valeur de cette constante peut être trouvée à l'aide de la distribution de χ^2 (khi-deux). L'orientation des axes principaux de l'ellipse est déterminée par la corrélation entre les variables. Cette ellipse joue le rôle d'une limite de contrôle pour le nuage de points dans le cas bidimensionnel et elle est nommée aussi ellipse de contrôle. Si un point dépasse la limite de contrôle elliptique on peut conclure que le processus est hors contrôle (une variable est hors contrôle ou la relation entre les variables s'est modifiée).

Quand le nombre de variables à contrôler - p - est supérieur à 2, la statistique T^2 de Hotelling s'écrit sous une forme matricielle de la manière suivante:

$$T^2 = (\bar{x} - \mu)^T \Sigma^{-1} (\bar{x} - \mu) \tag{1.8}$$

où μ est le vecteur des moyennes du processus de dimension $p \times 1$, Σ est la matrice de variance-covariance de dimension $p \times p$ et \bar{x} le vecteur des moyennes de l'échantillon.

Cette statistique est calculée à chaque instant et représentée sur une carte de contrôle qui ressemble à la carte de contrôle de Shewhart. Si l'on connaît les paramètres de processus, la limite de contrôle supérieure de la carte de contrôle T^2 de Hotelling est obtenue à l'aide de la distribution de khi-deux et la limite de contrôle inférieure est nulle:

$$LSC = \chi^2_{\left(\frac{\alpha}{2}, p\right)} \tag{1.9}$$

$$LIC = 0$$

où : $\chi^2_{(\alpha/2, p)}$ est le quantile d'ordre $\alpha/2$ de la distribution de khi-deux ayant p degrés de liberté.

Quand les paramètres du processus ne sont pas connus on est obligé de les estimer afin d'utiliser la carte de contrôle T^2 de Hotelling.

Estimation de μ et Σ

En pratique, il est nécessaire d'estimer le vecteur des moyennes du processus μ et la matrice de variance-covariance Σ à partir d'un groupe préliminaire d'échantillons de taille n , extrait d'un processus sous contrôle. Supposons qu'on dispose de m échantillons, les moyennes et les variances de chaque échantillon sont calculées comme suit :

$$\begin{aligned} \bar{x}_{jk} &= \frac{1}{n} \sum_{i=1}^n x_{ijk} \\ S_{jk}^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_{ijk} - \bar{x}_{jk})^2 \end{aligned}, \quad \begin{cases} j = 1, 2, \dots, p \\ k = 1, 2, \dots, m \end{cases} \quad (1.10)$$

où x_{ijk} est la i -ème observation appartenant à l'échantillon k effectuée sur la variable de processus j . La covariance entre la variable j et la variable h calculée à partir de l'échantillon k est :

$$S_{jkh} = \frac{1}{n-1} \sum_{i=1}^n (x_{ijk} - \bar{x}_{jk})(x_{ihk} - \bar{x}_{hk}), \quad \begin{cases} k = 1, 2, \dots, m \\ j \neq h \end{cases} \quad (1.11)$$

On calcule ensuite les valeurs moyennes des statistiques \bar{x}_{jk} , S_{jk}^2 et S_{jkh} :

$$\begin{aligned} \bar{\bar{x}}_j &= \frac{1}{m} \sum_{k=1}^m \bar{x}_{jk} \\ \bar{S}_j^2 &= \frac{1}{m} \sum_{k=1}^m S_{jk}^2 \end{aligned} \quad \text{où : } j = 1, 2, \dots, p$$

et

$$\bar{S}_{jh} = \frac{1}{m} \sum_{k=1}^m S_{jkh}, \quad \text{avec } j \neq h$$

Les valeurs $\bar{\bar{x}}_j$ sont des éléments constituant le vecteur estimé des moyennes du processus :

$$\bar{\bar{x}}^T = [\bar{\bar{x}}_1, \bar{\bar{x}}_2, \dots, \bar{\bar{x}}_p]. \quad (1.12)$$

L'expression de l'estimateur de la matrice de variance-covariance S est :

$$S = \begin{bmatrix} \bar{S}_1^2 & \bar{S}_{12} & \bar{S}_{13} & \cdots & \bar{S}_{1p} \\ & \bar{S}_2^2 & \bar{S}_{23} & \cdots & \bar{S}_{2p} \\ & & \bar{S}_3^2 & \vdots & \\ & & & \ddots & \vdots \\ & & & & \bar{S}_p^2 \end{bmatrix} \quad (1.13)$$

Si on remplace dans l'équation (1.8) les valeurs inconnues de la moyenne du processus et de la matrice de variance-covariance par leurs estimateurs calculés selon les équations (1.12) et (1.13), la statistique T^2 de Hotelling s'écrit :

$$T^2 = (\bar{x} - \bar{\bar{x}})^T S^{-1} (\bar{x} - \bar{\bar{x}}) \quad (1.14)$$

Alt [2] a démontré que dans ce cas la statistique T^2 de Hotelling est distribuée selon la distribution F de Fisher et les limites de contrôle pour l'étape I de la construction de la carte sont :

$$LSC = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{(\alpha, p, mn-m-p+1)} \quad (1.15)$$

$$LIC = 0$$

où p est le nombre de variables contrôlées; m est le nombre d'échantillons de taille n prélevés afin d'estimer les paramètres de processus et $F_{(\alpha, p, mn-m-p+1)}$ est le quantile d'ordre α de la distribution de Fisher avec p et $mn-m-p+1$ degrés de libertés.

A l'étape II de la construction de la carte, l'expression de la limite de contrôle supérieure devient :

$$LSC = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{(\alpha, p, mn-m-p+1)} \quad (1.16)$$

Si les paramètres du processus sont inconnus et la taille de l'échantillon est unitaire ($n=1$) la limite de contrôle supérieure de la carte de contrôle à l'étape II est obtenue à l'aide de la distribution F de Fisher selon la formule suivante :

$$LSC = \frac{p(m-1)(m+1)}{m(m-p)} F_{(\alpha, p, m-p)} \quad (1.17)$$

où $F_{(\alpha, p, m-p)}$ est le quantile d'ordre α de la distribution F avec p et $m-p$ degrés de liberté.

Quand le nombre des échantillons préliminaires m est grand ($m > 100$), l'approximation suivante pour la limite de contrôle supérieure de la carte T^2 de Hotelling est souvent utilisée :

$$LSC = \frac{p(m-1)}{m-p} F_{(\alpha, p, m-p)} \quad (1.18)$$

ou

$$LSC = \chi_{\alpha, p}^2 \quad (1.19)$$

Lowry et Montgomery [72] ont précisé que la limite khi-deux doit être utilisée avec précautions. En effet, si le nombre de variables est grand ($p > 10$) alors au moins 250 échantillons doivent être prélevés ($m > 250$) pour que la limite supérieure de contrôle de type khi-deux soit une approximation raisonnable de la valeur donnée par l'équation (1.17).

Tracy, Young et Mason [123] ont démontré que la distribution bêta doit être utilisée pour le calcul de la limite supérieure de contrôle à l'étape I si la taille de l'échantillon est unitaire :

$$LCS = \frac{(m-1)^2}{m} B\left(\alpha, \frac{p}{2}, \frac{m-p-1}{2}\right) \quad (1.20)$$

où $B\left(\alpha, \frac{p}{2}, \frac{m-p-1}{2}\right)$ est le quantile d'ordre α de la distribution bêta avec $p/2$ et $(m-p-1)/2$ comme paramètres.

En général, la limite inférieure de contrôle de la carte T^2 de Hotelling est fixée à zéro parce que cette statistique est toujours positive. Par conséquent, selon Tracy et al.[123], se rapporter à une limite inférieure de contrôle négative n'a pas de sens.

L'efficacité et les performances en détection de la carte de contrôle T^2 de Hotelling sont directement liées à la "qualité" de l'estimation des paramètres μ et Σ . Sullivan et Woodall [113] analysent cinq estimateurs de la matrice de variance-covariance dans le cas où la taille de l'échantillon est unitaire. Soit m observations indépendantes provenant d'une distribution p -normale :

$$x_i \sim N_p(\mu, \Sigma), \text{ où } i = 1, 2, \dots, m$$

Après l'arrangement des observations dans un vecteur colonne on a :

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_m^T \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,p} \end{bmatrix} \quad (1.21)$$

Le premier estimateur est la matrice de covariance des observations :

$$S_1 = \frac{1}{m-1} \sum_{i,j=1}^m (x_i - \bar{x})(x_j - \bar{x})^T \quad (1.22)$$

où m est le nombre d'observations, x_i est un vecteur d'observations de dimension $p \times 1$ et \bar{x} est le vecteur moyenne de toutes les observations.

Pour calculer le deuxième estimateur, les observations sont divisées en groupes de taille $p+1$. La matrice de variance-covariance de chaque groupe est calculée. L'estimateur S_2 de la matrice de variance-covariance du processus est calculé comme la moyenne des matrices variances-covariances des groupes.

Le troisième estimateur S_3 est similaire au deuxième sauf que des groupes superposés de taille r sont utilisés cette fois-ci. Le k -ème groupe contient les observations $k, k+1, \dots, k+r-1$. Les matrices de variance-covariance des groupes sont indépendantes du point de vue statistique dans le cas de l'estimateur S_2 mais pas pour l'estimateur S_3 .

Pour calculer le quatrième estimateur de la matrice de variance-covariance du processus, des groupes de taille 2 sont formés. Le vecteur différence des observations successives de chaque groupe est calculé :

$$y_i = x_{2i} - x_{2i-1}, \text{ où } i = 1, 2, \dots, \left\lfloor \frac{m}{2} \right\rfloor$$

où $\lfloor \cdot \rfloor$ représente l'arrondi à la valeur entière supérieure. L'estimateur de la matrice de variance-covariance Σ est :

$$S_4 = \frac{1}{2} \frac{Y^T Y}{\left\lfloor \frac{m}{2} \right\rfloor} \quad (1.23)$$

où la matrice Y est constituée des vecteurs lignes des différences.

Le cinquième estimateur suggéré par Holmes et Mergen [51] est calculé à l'aide de la différence entre des observations successives :

$$v_i = x_{i+1} - x_i, \quad \text{où : } i = 1, 2, \dots, m-1$$

L'estimateur de la matrice de variance-covariance Σ est :

$$S_5 = \frac{1}{2} \frac{V^T V}{(m-1)} \quad (1.24)$$

Scholz et Tosch [107] donnent une approximation de Wishart pour la distribution de S_5 . Wierda [128] démontre que les estimateurs S_4 et S_5 de la matrice de variance-covariance sont meilleurs que les estimateurs superposés (S_2 et S_3).

Les estimateurs S_1, \dots, S_5 sont des estimateurs sans biais. Enfin, pour la comparaison, le sixième estimateur est la valeur réelle de la matrice de variance-covariance :

$$S_6 = \Sigma \quad (1.25)$$

Sullivan et Woodall ont évalué par simulations les performances de la carte de contrôle T^2 de Hotelling dans la détection des causes assignables comme le saut en échelon ou la dérive en rampe en utilisant ces différents estimateurs de la matrice de variance-covariance.

Un saut en échelon dans la moyenne du processus à l'instant k peut être simulé ainsi :

$$\mu_i = \begin{cases} \mu_0, & i = 1, \dots, k \\ \mu_0 + \delta, & i = k + 1, \dots, m \end{cases}$$

où μ_0 est le vecteur des moyennes sous-contrôle, et δ est l'amplitude du dérèglement.

La détection du saut à l'aide de la carte de contrôle T^2 basée sur les estimateurs S_2 et S_4 est fonction de l'instant d'apparition du saut. Un saut situé à l'extrémité d'un groupe n'affecte pas l'estimation de la matrice de variance-covariance Σ .

La carte T^2 de Hotelling basée sur l'estimateur S_1 , utilisée habituellement dans la pratique, est insensible à l'amplitude du saut, donc cet estimateur est inadéquat pour détecter des sauts de moyenne. Ses performances dans la détection d'une situation hors-contrôle s'améliorent si le saut est situé au début ou à la fin des observations.

Les deux estimateurs qui utilisent les groupes superposés, S_3 et S_5 , donnent de meilleurs résultats que les estimateurs S_2 et S_4 .

En ce qui concerne une dérive en rampe dans le vecteur des moyennes qui est décrite par l'équation suivante :

$$\mu_i = \mu_0 + \frac{i-1}{m-1} \delta, i = 1, \dots, m$$

où μ_0 est le vecteur des moyennes sous contrôle et δ est le vecteur différence entre le dernier et le premier vecteur des moyennes la carte T^2 basée sur l'estimateur S_1 reste toujours moins performante que la carte obtenue en utilisant les autres estimateurs.

Comme dans le cas d'un saut en échelon, la carte T^2 basée sur les estimateurs S_3 et S_5 est plus performante que la carte T^2 basée sur les estimateurs S_2 et S_4 . Sullivan et Woodall ont montré que la carte T^2 construite avec les estimateurs S_1, \dots, S_5 présente des performances modestes dans la détection des valeurs aberrantes. Quand le nombre de valeurs aberrantes augmente, la matrice de variance-covariance Σ est surestimée et donc la probabilité de détection d'une situation hors-contrôle est faible.

Les cartes de contrôle MEWMA et MCUSUM

La carte de contrôle T^2 de Hotelling présentée dans le précédent paragraphe utilise seulement le vecteur d'observations à l'instant présent pour déterminer l'état du processus et elle est donc moins performante pour la détection des faibles dérèglages du processus. Pour remédier à cet inconvénient, les cartes de contrôle MEWMA (Multivariate EWMA) et MCUSUM (Multivariate CUSUM) ont été développées en analogie avec le cas univarié.

Les cartes de contrôle multivariées sont de deux types : *non-directionnelles* ou *directionnelles* (voir Lowry et Montgomery [72]). La POM des cartes de contrôle non-directionnelles est déterminée uniquement en fonction de la distance entre la moyenne de l'échantillon hors-contrôle et la moyenne du processus sous-contrôle. La POM des cartes de contrôle directionnelles est en fonction de cette distance et de la direction dans laquelle se trouve, dans l'espace multidimensionnel, la moyenne de l'échantillon hors-contrôle par rapport à la moyenne du processus sous-contrôle. Les cartes de contrôle EWMA et CUSUM univariées sont non-directionnelles, mais leurs variantes multidimensionnelles ne le sont pas.

La construction de la carte MEWMA (Lowry et al., [73]) repose sur le même principe que la carte univariée EWMA. Les équations pour la carte MEWMA sous forme matricielle sont présentées ci-dessous :

$$Y_k = R X_k + (I - R) Y_{k-1} \quad (1.26)$$

où X_k est le vecteur d'observations, R est la matrice diagonale de pondération des variables et I est la matrice identité.

La valeur de départ Y_0 dans l'équation (1.26) est souvent considérée égale au vecteur cible du processus. Si on donne le même poids à toutes les variables dans le calcul de la statistique Y_k la matrice diagonale de pondération R devient une constante de pondération r ayant une valeur comprise entre 0 et 1. L'équation (1.26) s'écrit dans ce cas :

$$Y_k = rX_k + (1 - r)Y_{k-1} \quad (1.27)$$

Le processus est considéré comme hors-contrôle à chaque fois que la statistique T_i^2 dans l'équation (1.28) dépasse une certaine limite ($h_M > 0$) calculée pour respecter une POM_0 (Période Opérationnelle Moyenne quand le processus est sous contrôle) :

$$T_i^2 = Y_k^T \Sigma_{Y_k}^{-1} Y_k > h_M \quad (1.28)$$

où Σ_{Y_k} est la matrice de variance-covariance de la statistique Y_k . Elle est calculée à l'aide de la formule :

$$\Sigma_{Y_k} = \frac{r[1 - (1 - r)^{2k}]}{2 - r} \Sigma \quad (1.29)$$

où Σ est la matrice de variance-covariance pour les données originales X_k .

La valeur optimale pour r est fonction du nombre de variables contrôlées et de l'amplitude du dérèglement que l'on souhaite détecter. En règle générale la valeur de r sera choisie d'autant plus petite que l'amplitude du dérèglement qu'on souhaite détecter sera faible. Pour la carte MEWMA c'est par le choix de r que certaines directions peuvent être favorisées. Si on donne le même poids à toutes les variables alors la carte est non-directionnelle. Si une situation hors-contrôle est mise en évidence par une carte MEWMA avec des pondérations différentes pour les variables, elle est attribuée aux variables qui ont les poids les plus importants.

Prahu et Runger [98] ont calculé la POM de la carte MEWMA à l'aide des chaînes de Markov bidimensionnelles. Ils ont abouti aux mêmes conclusions que Lowry [73] pour la carte EWMA unidimensionnelle, c'est-à-dire que la carte de contrôle MEWMA donne de meilleurs résultats dans la détection des faibles dérèglements par rapport à la carte T^2 de Hotelling.

La carte MCUSUM est une extension de la carte CUSUM univariée pour le cas multidimensionnel. Une des meilleures cartes MCUSUM pour détecter des dérèglements de la moyenne du processus a été développée par Pignatiello et Runger [94]. Elle est construite à partir du vecteur :

$$C_t = \sum_{i=t-n_t+1}^t (X_i - \mu_0) \quad (1.30)$$

qui représente la somme cumulée des écarts entre le vecteur d'observations X_i et le vecteur des moyennes sous-contrôle μ_0 . Ensuite, la norme du vecteur C_t est calculée selon la formule :

$$\|C_t\| = \sqrt{C_t^T \Sigma^{-1} C_t} \quad (1.31)$$

La statistique de contrôle est :

$$MC1_t = \max(\|C_t\| - kn_t, 0) \quad (1.32)$$

où :

$$n_t = \begin{cases} n_{t-1} + 1 & \text{si } MC1_{t-1} > 0 \\ 1 & \text{si non} \end{cases} \quad \text{et } \mathbf{k} \text{ est un paramètre qui dépend de l'amplitude du dérèglement qu'on veut détecter.}$$

Les cartes MCUSUM sont plutôt non-directionnelles, donc elles sont dépendantes seulement du paramètre de non-centralité (Crosier, [17]).

Les cartes MEWMA et MCUSUM présentent une certaine inertie (retard) dans la détection d'un dérèglement si celui-ci se produit dans une direction opposée à l'évolution du processus. Pour remédier à cet inconvénient des limites de type Hotelling peuvent être ajoutées sur la carte afin de faciliter la détection (Lowry et Montgomery, [72]).

1.6 Conclusion

La "Qualité" est devenu un des mots clés figurant dans la stratégie des entreprises d'aujourd'hui. L'évolution vers la "Qualité", telle qu'on la connaît actuellement, a été lente mais constante au fil du temps, voir l'annexe A.1. Nous avons analysé en bref, dans ce chapitre introductif, seulement un aspect particulier de la "Qualité" – l'aspect lié à l'utilisation des techniques statistiques pour la maîtrise des processus. Nous avons mentionné la place occupée par la maîtrise statistique de processus (MSP) parmi les outils du contrôle statistique de la qualité (CSQ) et nous avons présenté des notions, très générales, concernant les cartes de contrôle. Le rôle joué par les cartes de contrôle univariées dans la démarche d'amélioration de la qualité des processus industriels et dans le changement de la mentalité par rapport à l'application des méthodes statistiques dans l'industrie a été commenté. Finalement, la nécessité de l'application des techniques relevant du contrôle statistique multivarié afin d'assurer la maîtrise des processus complexes a été évoquée. Nous avons présenté, d'une manière succincte, quelques unes des cartes de contrôle multivariées les plus utilisées. En dépit de leurs avantages par rapport aux cartes de contrôle univariées, un problème très important qui reste à résoudre est la manière d'identifier la variable ou les variables responsable(s) d'une situation hors-contrôle. De nombreuses recherches ont été effectuées ces dernières années pour remédier à cet inconvénient. Nous allons présenter dans le chapitre suivant quelques solutions, parmi les plus représentatives, proposées pour améliorer la détection de la cause d'une situation hors-contrôle dans le cas multivarié.

Chapitre 2

- 2.1 Introduction
- 2.2 Panorama des méthodes de détection
- 2.3 Description et analyse des méthodes existantes
- 2.4 Conclusion

2.1 Introduction

Les cartes de contrôle statistique sont des outils graphiques nous permettant de visualiser l'état d'un processus dans le temps. Depuis l'apparition de la carte de contrôle \bar{X}, R (moyenne, étendue) de *Shewhart*, dans les années 30, les cartes de contrôle n'ont pas cessé d'évoluer, gardant pourtant leur caractère univarié. Même si le progrès enregistré en utilisant les cartes de contrôle univariées est indiscutable, la réalité multivariée de tous les processus industriels, a imposé l'apparition de nouvelles méthodes basées sur la statistique multivariée. Les approches théoriques existaient déjà mais la difficulté de les mettre en place les a fait oublier pendant une vingtaine d'années. Aujourd'hui, avec l'avènement de l'ordinateur et de l'électronique, les méthodes jadis abandonnées ont été reconsidérées.

Les cartes de contrôle multivariées ont l'avantage de prendre en compte les corrélations existantes entre les variables. Parmi les cartes de contrôle multivariées, la carte T^2 de Hotelling [53] est peut être la plus connue et la plus utilisée. Les techniques utilisées pour la construction des cartes univariées CUSUM et EWMA ont été adaptées pour les situations multivariées donnant lieu à l'apparition des cartes multivariées correspondantes MCUSUM et MEWMA. Nous avons présenté les principes de construction de ces cartes au chapitre 1.

L'inconvénient majeur des cartes de contrôle multivariées est qu'elles ne nous donnent aucune indication concernant la variable, ou les variables, responsable(s) de l'apparition d'une situation hors contrôle. Ainsi, il est difficile de prendre les actions correctives nécessaires afin d'améliorer le processus. Cela reste un problème difficile même dans le cas univarié car parfois il n'y a pas une relation directe entre la variable suivie et les actions correctives. Par exemple, supposons qu'on effectue des mesures sur le

rétrécissement d'une pièce issue d'un processus de moulage par injection. Si une alerte est donnée sur la carte de contrôle, il n'y a pas un moyen direct de réglage pour corriger la situation. Par contre, il existe une large palette de possibilités indirectes, comme le contrôle de la pression, de la température, du flux, et du temps afin de remédier au problème. Un changement d'un de ces facteurs peut influencer directement le rétrécissement. Déterminer la cause primaire d'un problème nécessite des connaissances approfondies sur le processus et des investigations minutieuses.

La détermination de la cause d'un dérèglement dans le cas multivarié est encore plus difficile. Des recherches plus élaborées et plus soutenues doivent être menées encore dans cette direction.

Dans les prochains paragraphes nous allons passer en revue et commenter quelques techniques actuellement employées pour la détermination de la source d'un dérèglement dans le cas d'un processus multivarié.

2.2 Panorama des principes de détection

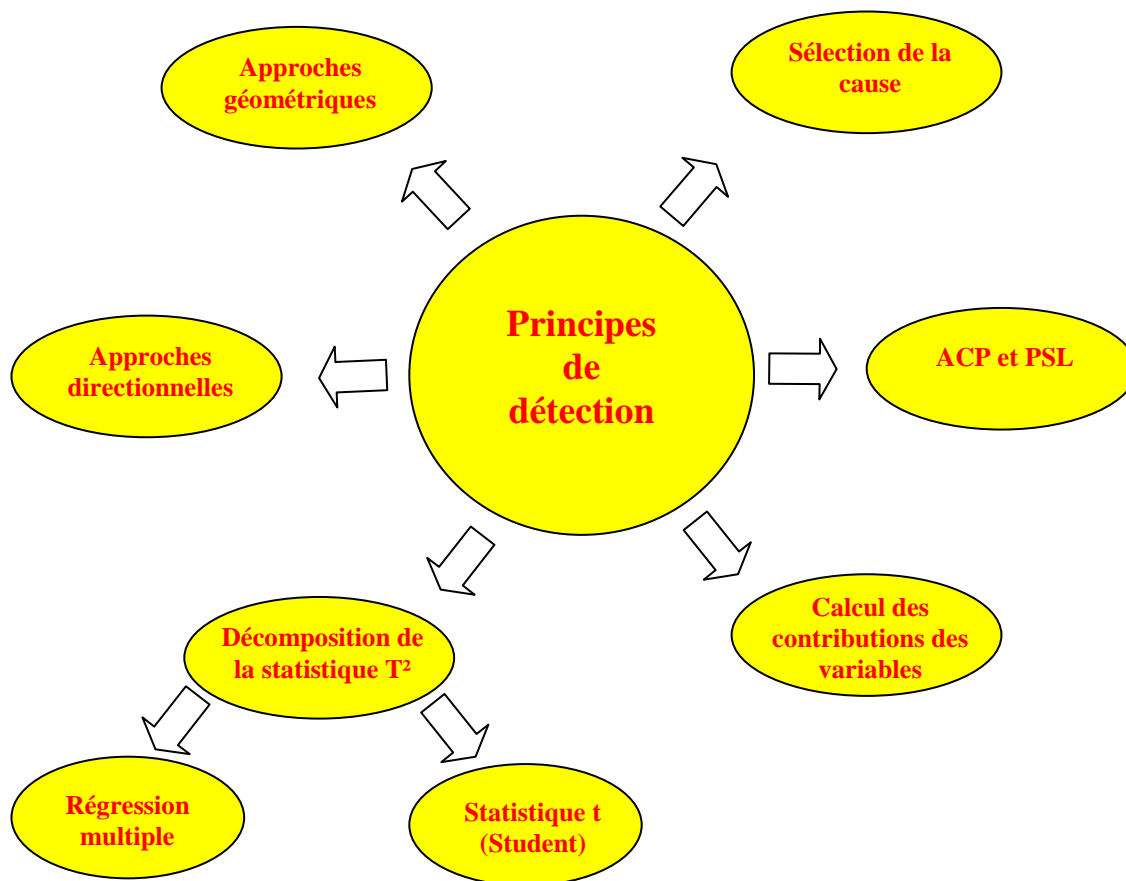


figure II.1 – Panorama des principes de détection

Comme on l'a précisé dans la partie introductive, des cartes de contrôle plus ou moins performantes ont été développées depuis une vingtaine d'années, afin de détecter les situations hors contrôle dans les processus multivariés. Parmi ces cartes, la carte T^2 de Hotelling est, sans doute, la plus utilisée. Malheureusement, la carte de contrôle Hotelling ou bien les autres cartes multivariées comme MCUSUM ou MEWMA, même si elles sont assez performantes pour détecter les situations hors contrôle, ne nous donnent pas d'informations concernant la variable responsable d'un dérèglement. Par conséquent, l'opérateur est obligé de procéder par des tâtonnements pour identifier la source du problème ce qui n'est pas toujours très facile et peut conduire à l'augmentation de la variabilité et de l'instabilité du processus. Beaucoup de recherches ont été effectuées pour localiser le dérèglement sur une carte multivariée. Un résumé non-exhaustif des principes de détection est donné dans la figure II.1.

Nous allons, dans le paragraphe suivant, donner une description sommaire de ces méthodes et mentionner leurs avantages et leurs inconvénients.

2.3 Description et analyse des méthodes existantes

2.3.1 Méthode de régression des variables

Hawkins [46] propose l'ajustement des variables par la régression multiple pour déterminer la source d'une situation hors contrôle dans le cas multivarié. Il continue ainsi le travail de Mandel [76] qui a démontré que le suivi, à l'aide d'une carte de contrôle classique, des résidus obtenus par la régression d'une variable quelconque dans les autres donne de meilleurs résultats que le suivi de la variable elle-même.

Hawkins prend en compte deux mécanismes de dérèglement dans un processus:

- ✓ le mécanisme où certaines variables peuvent être dépendantes l'une de l'autre mais, où une modification sur l'une n'affecte pas les autres. Un exemple, de ce type, est le processus de raffinage de platine. Six métaux entrant dans la composition d'alliages à base de platine ont différentes concentrations dans des matériaux intermédiaires. Ces concentrations ont tendance à être corrélées. Dans le laboratoire de test, il peut arriver de noter d'une façon erronée la quantité d'un de ces métaux. Evidemment, ce biais ne va pas affecter les tests effectués sur les autres métaux. Il semble très logique qu'un problème au niveau d'un métal quelconque puisse être détecté plus rapidement si on utilise judicieusement les résultats des mesures sur les autres métaux.
- ✓ le mécanisme où les variables se trouvent dans une structure hiérarchique telle que la modification apparue dans une certaine variable peut affecter toutes les variables qui suivent mais n'affecte pas les variables qui la précèdent. Ce type de processus est appelé processus en cascade. Un processus métallurgique où les matériaux passent par différentes étapes d'enrichissement est un bon exemple pour ce type d'association. La qualité du produit à la fin d'une étape dépend de la réussite des opérations effectuées à cette étape mais aussi de ce qui s'est passé aux étapes antérieures. Si une étape n'a pas été accomplie correctement elle peut

affecter les variables du processus non seulement à cette étape mais aussi pendant toutes les étapes ultérieures. Cela peut présenter l'inconvénient de considérer, à tort, toutes les mesures faites ultérieurement comme hors contrôle.

Soit X_i ($p \times 1$) le vecteur d'observations à l'instant i sur p variables de processus. Si le processus est sous contrôle, on considère que les observations X_i sont indépendantes et distribuées selon une distribution normale multivariée de moyenne μ et de matrice de variance-covariance Σ . La valeur de la variable j à l'instant i est notée X_{ij} . Les valeurs de la variable j prises à des instants non-spécifiés sont notées X_j .

Si on suppose que le vecteur des moyennes μ d'un processus p -varié (p variables) prend une certaine valeur μ_A dans le cas d'un saut de la moyenne du processus, le test optimal selon Healy [50] pour détecter le saut est:

$$(\mu_A - \mu)^T \Sigma^{-1} (X_i - \mu) = (\mu_A - \mu)^T Z_i \quad (2.1)$$

où : $Z_i = \Sigma^{-1} (X_i - \mu)$ est un vecteur colonne de dimension $p \times 1$

X_i est le vecteur d'observations des variables de processus à l'instant i

μ est le vecteur des moyennes

Σ est la matrice de variance-covariance

Hawkins [47] combine les résultats de Mandel et Healy et affirme que si la différence entre le vecteur μ_A et le vecteur μ est due à un saut de moyenne dans la variable j , alors, pour détecter ce saut, il suffit de regarder Z_j la j -ème composante du vecteur Z . Le vecteur Z est nommé vecteur d'ajustement par régression des variables et sa j -ème composante Z_j représente, en effet, le résidu standardisé de la régression de la variable X_j dans toutes les autres variables. En pratique, pour contrôler le processus, il faut utiliser p cartes de contrôle pour les résidus Z_j , c'est-à-dire une pour chaque variable de processus. Si sur la carte de contrôle Z_j on a une alerte, alors on peut dire que c'est la variable X_j qui a subi un dérèglement.

L'approche de Hawkins donne aussi des bons résultats dans le cas particulier des processus en cascade. Dans ce cas précis, la régression de la variable X_j , qui a subi le saut de moyenne, ne se fait pas en prenant toutes les autres variables comme exogènes, mais en prenant en compte seulement les variables $X_1 \dots X_{j-1}$. Soit Y_j le résidu standardisé qui est obtenu après avoir explicité la variable X_j à l'aide des variables $X_1 \dots X_{j-1}$. Quand le processus est sous contrôle, la variable Y_j est distribuée selon une distribution normale centrée réduite $N(0,1)$. Un changement de la moyenne de X_j va affecter aussi les variables $X_{j+1} \dots X_p$ et donc il y aura un changement dans la distribution de Y_j . En pratique, on doit observer simultanément des diagrammes de contrôle Y_j pour toutes les variables afin de détecter la source d'un saut dans le cas multivarié.

L'avantage d'utiliser les variables Y et Z pour le contrôle est la réduction de la variabilité inhérente au processus et par conséquent la mise en évidence plus nette du saut. En effet, les résidus de la variable X_j ont une variance réduite par le facteur : $\sqrt{1 - R_j^2}$ (où : R_j – est le coefficient de corrélation multiple) par rapport à la variance de la variable X_j . Utiliser les résidus permet d'éliminer l'inconvénient lié à la corrélation entre les variables car les résidus sont indépendants et identiquement distribués. On peut donc tracer pour les résidus des cartes de contrôles classiques de type Shewhart, CUSUM ou EWMA sans avoir de problèmes pour le calcul des limites de contrôle.

Evidemment pour appliquer la méthode proposée par Hawkins et bien choisir entre les cartes de contrôle \mathbf{Y} et \mathbf{Z} , il est nécessaire d'avoir de bonnes connaissances concernant le processus. Ces connaissances s'obtiennent soit en demandant l'avis des experts, soit par une très bonne analyse exploratoire des données. Parfois, néanmoins, il est difficile de hiérarchiser les variables comme le préconise Hawkins dans sa démarche, surtout pour des systèmes complexes. Il a été démontré par Hawkins que dans le cas où cette hiérarchisation des variables n'est pas correctement choisie, les résultats obtenus par la méthode de régression des variables sont inférieurs aux résultats qu'on aurait pu obtenir en observant les variables directement à l'aide d'une carte de contrôle classique. Un autre inconvénient de la méthode est le grand nombre de cartes que l'opérateur doit surveiller en même temps (au moins une carte de contrôle par variable) ce qui n'est pas toujours très facile en pratique. Finalement, il faut préciser qu'il est difficile de déterminer la source du dérèglement à l'aide de cette méthode quand plusieurs variables de processus changent simultanément.

2.3.2 Méthode basée sur l'utilisation de la statistique t de Student

Doganaksoy, Faltin et Tucker [25] proposent une démarche basée sur l'utilisation de la statistique t (*Student*) afin de déterminer la source d'un dérèglement sur une carte de contrôle multivariée (la carte de contrôle T^2 de Hotelling). Leur approche se décompose en deux étapes. Dans la première, les variables sont classées selon leur importance en calculant leurs contributions individuelles au dérèglement et après, dans une deuxième étape, est déterminée la variable qui a subi un changement ou éventuellement les variables qui ont subi un changement.

Dès qu'une situation hors contrôle a été détectée sur la carte de contrôle T^2 , les variables sont groupées en ordre décroissant de l'importance que chacune a dans la construction de la carte T^2 en utilisant la statistique univariée t (Student) comme indicateur :

$$t = \frac{(\bar{x}_i^n - \bar{x}_i^r)}{\sqrt{s_{ii} \left(\frac{1}{n^n} + \frac{1}{n^r} \right)}} \quad (2.2)$$

où : \bar{x}_i^n est le nouvel échantillon de la variable i

\bar{x}_i^r est la moyenne de référence de la variable i

s_{ii} est l'élément (i, i) de la matrice \mathbf{S} de variance-covariance

n^n est la taille du nouvel échantillon

n^r est la taille de l'échantillon de référence

On obtient ainsi une estimation relative sur les variables les plus probablement responsables du dérèglement. En pratique, après avoir groupé les variables en ordre décroissant de leur contribution à la statistique multivariée, il est parfois possible d'identifier la source de l'anomalie par une bonne interprétation du processus de mise en oeuvre. Par exemple, si plusieurs caractéristiques d'une pièce présentent des valeurs importantes de la statistique t , et sachant que ces caractéristiques sont le résultat des opérations réalisées sur la même machine, la cause du problème est évidente (la machine).

Il ne faut pas oublier pourtant que cette méthode d'arrangement des variables nous donne seulement une indication relative sur les variables probablement responsables du dérèglement. La variable k , par exemple, est considérée comme très probablement responsable du dérèglement, si l'intervalle de confiance pour la différence $\mu_k^r - \mu_k^n$ ne contient pas la valeur zéro. La construction appropriée de l'intervalle de confiance pour $\mu_i^r - \mu_i^n$, où $i = 1 \dots p$, devient un problème très important. Pour décider quelle variable ou groupe de variables est la source d'un dérèglement sur la carte T^2 Doganaksoy, Faltin et Tucker proposent l'utilisation des limites de contrôle de type Bonferroni.

La méthode est mise en pratique en cinq étapes (figure II.2):

- 1) Utiliser la carte de contrôle multivariée T^2 de Hotelling avec la limite de contrôle fixée pour un niveau de confiance nominal α_{nom} . Si la carte T^2 détecte une situation hors contrôle passer à l'étape 2;
- 2) Pour chaque variable dans le vecteur d'observations calculer la statistique t et la fonction de répartition cumulée $T(t,d)$ de t , où d est le nombre de degrés de liberté de t . Calculer pour chaque variable :

$$\alpha_i = |2T(t; n^r - 1) - 1|, \text{ où } i = 1 \dots p$$

le plus petit niveau de confiance qui conduit à un intervalle de confiance pour $\mu_i^r - \mu_i^n$ contenant la valeur nulle.

- 3) Afficher les valeurs α_i sur un diagramme. Les variables avec les plus larges valeurs α_i sont les variables ayant les plus larges valeurs de la statistique univariée t et donc probablement responsables du dérèglement.
- 4) Calculer le niveau de confiance α_{Bonf} de la différence $\mu_i^r - \mu_i^n$ pour $i = 1 \dots p$, avec la formule:

$$\alpha_{Bonf} = \frac{p + \alpha_{sim} - 1}{p}$$
 où: α_{sim} est le niveau de confiance désiré pour l'intervalle de confiance de type Bonferroni pour $\mu_i^r - \mu_i^n$.
- 5) Les variables ayant $\alpha_i > \alpha_{Bonf}$ sont retenues comme celles qui sont les plus probablement responsables du dérèglement.

Dans de nombreuses situations, les intervalles de confiance de type Bonferroni ne mettent pas en évidence le dérèglement même si la carte de contrôle T^2 de Hotelling le détecte. Pour augmenter la fréquence de détection sur les variables individuelles Doganaksoy, Faltin et Tucker ont utilisé un niveau de confiance pour la limite de type Bonferroni inférieur à α_{nom} . Pourtant, ce niveau de confiance α_{sim} ne doit pas être très réduit afin de ne pas augmenter le nombre de fausses alarmes. Le choix optimal du niveau de confiance α_{sim} doit se faire en fonction du nombre de variables, de la corrélation entre les variables, du risque α_{nom} de la carte T^2 , du nombre accepté de fausses alarmes, etc... Doganaksoy, Faltin et Tucker ont déterminé à partir des simulations de type Monte Carlo que la valeur optimale pour α_{sim} doit être à peu près égale à 0.8.

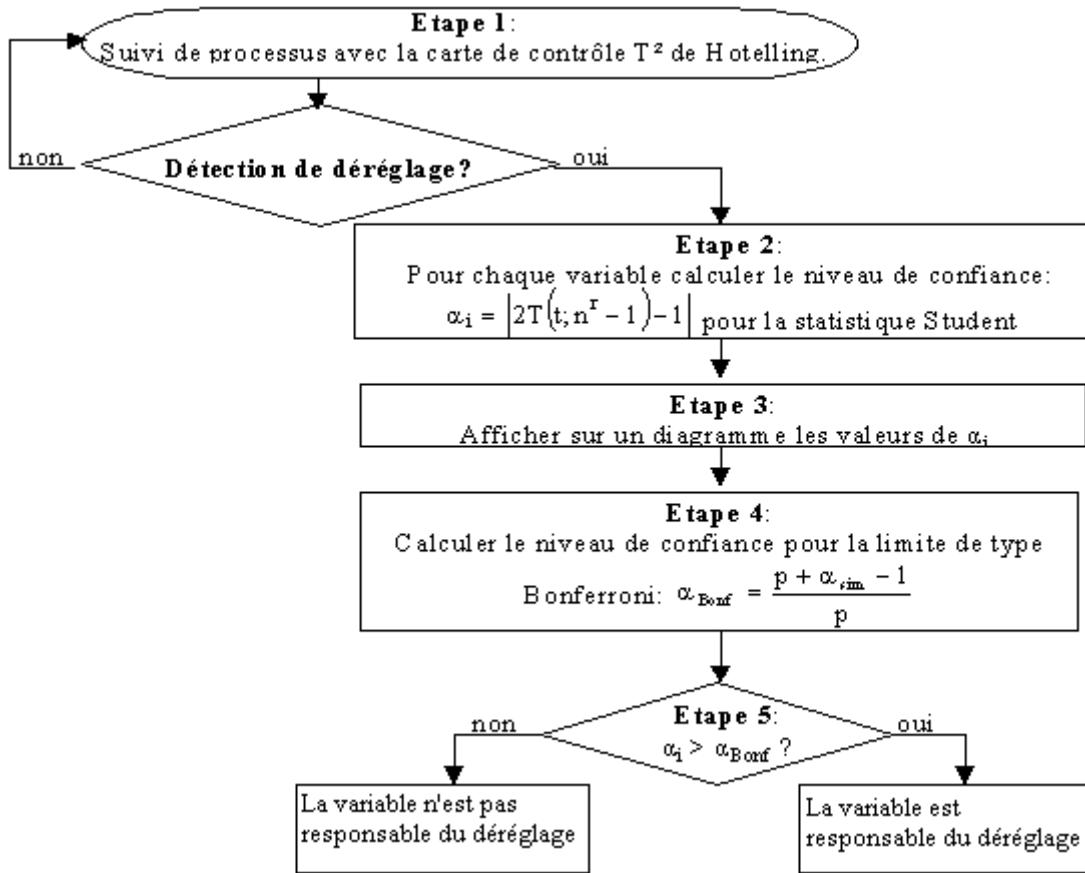


figure II.2 - Schéma de contrôle à l'aide de la statistique t (Student)

Un des inconvénients de la méthode réside dans la détermination des intervalles de confiance de type Bonferroni. Réduire la taille de la région de type Bonferroni pour augmenter la probabilité de détection d'un dérèglement, peut conduire à l'augmentation des fausses alertes. En utilisant les intervalles simultanés Bonferroni en combinaison avec la méthode d'arrangement des variables selon la valeur de la statistique student t on obtient néanmoins un outil plus robuste dans la détection de la variable responsable d'un dérèglement. Un autre inconvénient majeur de la méthode est que la corrélation entre les variables n'est pas prise en compte dans les calculs car la statistique student utilisée est par sa nature univariée. Cette méthode peut donner de bons résultats uniquement dans le cas où les variables sont indépendantes.

2.3.3 Décomposition MYT de la carte de contrôle T^2 de Hotelling

Cette méthode a été proposée par Mason, Young et Tracy [80] et elle est connue aussi sous le nom de décomposition MYT. Il est possible de démontrer que les deux méthodes présentées dans les paragraphes antérieurs (la méthode de régression proposée par Hawkins - paragraphe 2.3.1 - et la méthode basée sur l'utilisation de la statistique t de Student proposée par Doganaksoy, Faltin et Tucker - paragraphe 2.3.2) sont des cas particuliers de la méthode de décomposition MYT.

La difficulté liée à la détection de la cause d'un dérèglement est l'inconvénient majeur de la statistique T^2 de Hotelling, dont la formule est présentée ci-dessous :

$$T^2 = (\mathbf{X}_i - \bar{\mathbf{X}})^T S^{-1} (\mathbf{X}_i - \bar{\mathbf{X}}) \quad (2.3)$$

où $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ip})^T$ représente le vecteur d'observations à l'instant i , $\bar{\mathbf{X}}$ l'estimateur du vecteur des moyennes et S^{-1} l'estimateur de la matrice variance-covariance;

Il sera utile, dans ce cas, de déterminer l'effet réel de chaque variable dans le calcul de la statistique T^2 . De telles connaissances pourront aider à interpréter les situations hors contrôle mises en évidence sur la carte.

La méthode proposée par Mason, Tracy et Young consiste à décomposer la statistique T^2 dans un nombre limité de composantes orthogonales qui à leur tour sont aussi des distances statistiques:

$$T^2 = T_1^2 + T_{2 \bullet 1}^2 + T_{3 \bullet 1,2}^2 + T_{4 \bullet 1,2,3}^2 + \dots + T_{p \bullet 1,2,\dots,p-1}^2 = T_1^2 + \sum_{j=1}^{p-1} T_{(j+1) \bullet 1,2,\dots,j}^2 \quad (2.4)$$

où $T_1^2 = \frac{(\mathbf{X}_{i1} - \bar{X}_1)^2}{s_1^2}$ est la statistique univariée t de Student de la variable \mathbf{X}_1 élevée au carré;

Cette statistique a été utilisée par Doganaksoy, Faltin et Tucker dans la méthode décrite au paragraphe 2.3.2, voir l'équation (2.2). On remarque dans l'équation (2.4) que le terme T_1^2 représente seulement un des nombreux termes de la décomposition de la statistique T^2 (le terme non-conditionné);

$$T_{p \bullet 1,2,\dots,p-1} = \frac{X_{ip} - \bar{X}_{p \bullet 1,2,\dots,p-1}}{s_{p \bullet 1,2,\dots,p-1}}$$

est la p -ème composante du vecteur d'observations \mathbf{X}_i centrée et réduite par rapport à l'estimateur de la moyenne et de l'écart-type de la distribution conditionnelle de X_p sachant X_1, X_2, \dots, X_{p-1} , avec:

$$\bar{X}_{p \bullet 1,2,\dots,p-1} = \bar{X}_p + \mathbf{b}_p^T (\mathbf{X}_i^{(p-1)} - \bar{\mathbf{X}}^{(p-1)})$$

où : \bar{X}_p est l'estimateur sur n observations de la moyenne de la variable X_p

$\mathbf{b}_p = S_{XX}^{-1} s_{Xp}$ est un vecteur de dimension $(p-1)$ qui donne les estimateurs des coefficients de régression de la variable X_p dans les premières $p-1$ variables.

$$s_{p \bullet 1,2,\dots,p-1}^2 = s_x^2 - s_{XX}^T S_{XX}^{-1} s_{Xp}$$

et

$$S = \begin{bmatrix} S_{XX} & s_{Xp} \\ s_{Xp}^T & s_x^2 \end{bmatrix}$$

Dans la décomposition MYT, voir l'équation (2.4), on peut distinguer deux catégories de facteurs, les facteurs non-conditionnés (ne dépendent pas des autres

variables, par exemple T_1^2) et les facteurs conditionnés (dépendent des autres variables, par exemple $T_{2,1}^2$, etc..). Il est intéressant de remarquer que la décomposition n'est pas unique. Il existe en réalité $p!$ différentes manières possibles de décomposition de la statistique T^2 . Par exemple, nous pouvons commencer la décomposition par n'importe quelle variable parmi les p variables. Ensuite, nous pouvons choisir n'importe quelle autre variable parmi les $p-1$ restantes pour constituer la première composante conditionnée et ainsi de suite.

Dans l'équation (2.4) tous les p termes sont positifs et contribuent à l'augmentation de la statistique globale T^2 . Si la statistique globale T^2 met en évidence une situation hors contrôle, alors il sera possible de déterminer quelle composante, ou quelles composantes, apporte(ent) le plus de contribution à l'augmentation de la statistique. Ce constat représente la raison principale en faveur de la décomposition de la statistique T^2 . La question qui se pose maintenant est comment peut-on utiliser l'information contenue dans chacune de ces p composantes de la décomposition ?

Premièrement, on doit observer que les p termes de la décomposition de T^2 , voir l'équation (2.4), sont indépendants. Deuxièmement, chacun de ces termes est distribué selon une distribution F de Fisher ayant 1 et $n-1$ degrés de liberté:

$$T_{j+1,1,\dots,j}^2 \sim \frac{n+1}{n} F_{(1,n-1)} \quad (2.5)$$

et on peut tester chaque terme, à l'aide du test de Fisher afin de savoir s'il est statistiquement significatif .

Si dans l'équation (2.3) on remplace μ et Σ par leurs estimateurs \bar{X} et S , après quelques transformations mathématiques, il est possible de démontrer que la j -ème composante du vecteur Z , dans l'équation (2.1) prend la forme:

$$Z_j = \frac{T_{j,1,\dots,j-1,j+1,\dots,p}}{S_{j,1,\dots,j-1,j+1,\dots,p}} \quad (2.6)$$

Hawkins a montré que Z_j représente le résidu standardisé issu de la régression de la variable X_j dans toutes les autres variables du vecteur d'observations X . En plus, Z_j représente, à part un terme constant (l'inverse de l'écart-type conditionné), la racine carrée d'un des termes conditionnés de la décomposition de la statistique T^2 . Donc, la statistique de Hawkins basée sur la régression est un élément de diagnostic dans l'interprétation d'une situation hors contrôle à l'aide de la carte T^2 de Hotelling. Pourtant, comme on peut remarquer dans l'équation (2.4) celui-ci n'est qu'un des termes conditionnés de la décomposition de la statistique T^2 .

Dans une deuxième démarche, Hawkins utilise la régression de la j -ème variable uniquement dans les variables $1,2,\dots,j-1$ qui la précèdent. Selon les notations utilisées dans la décomposition de la statistique T^2 de Hotelling, la j -ème composante du vecteur Y est donnée par:

$$Y_j = \frac{T_{j,1,\dots,j-1}}{S_{j,1,\dots,j-1}} \quad (2.7)$$

donc, Y_j est aussi lié à un terme conditionné de la décomposition de T^2 .

Les facteurs non-conditionnés nous signalent si une variable de processus a subi un dérèglement. Les facteurs conditionnés nous aident à déterminer si la relation entre les variables s'est modifiée (figure II.3). Le calcul des facteurs conditionnés est fait par la régression multiple d'une variable dans toutes les variables qui la précèdent.

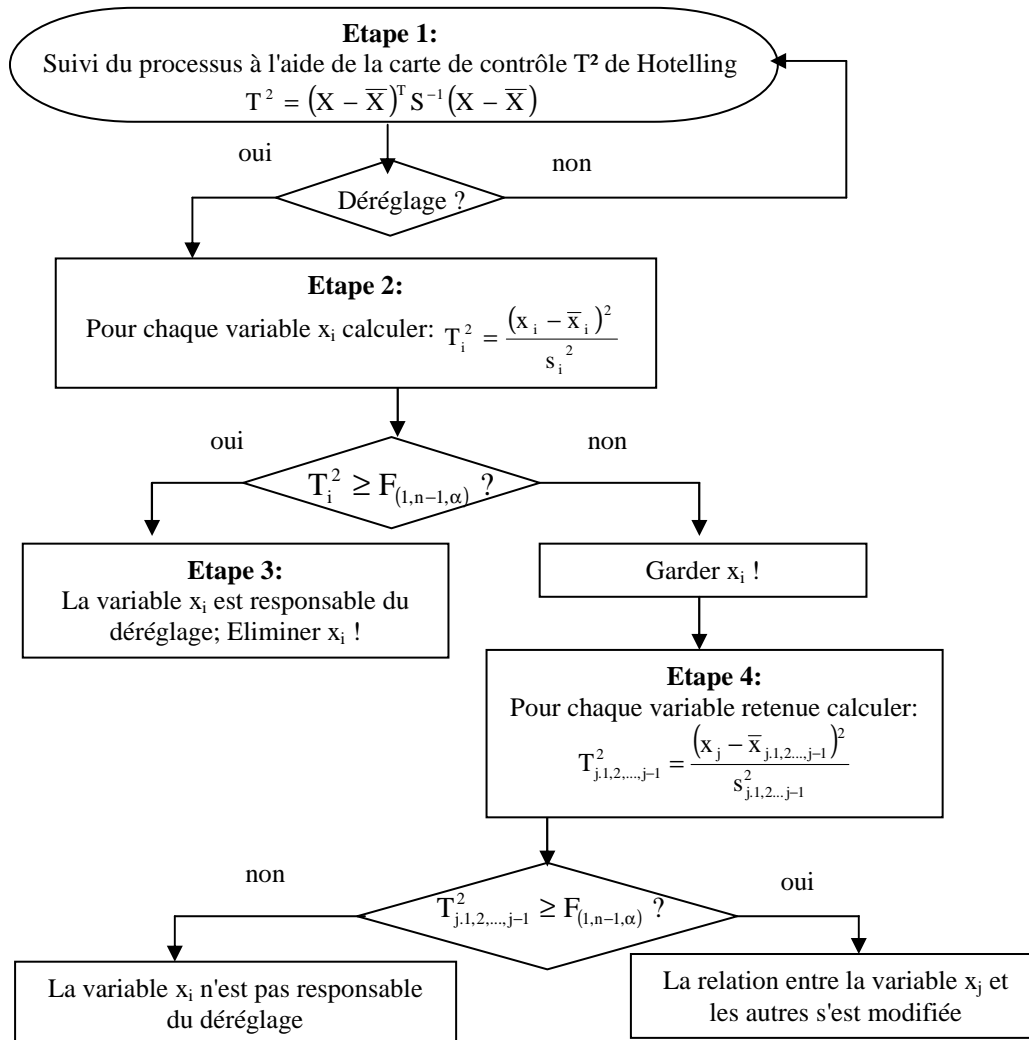


figure II.3 - Décomposition MYT de la statistique T^2 de Hotelling

L'avantage majeur de la méthode de décomposition de la statistique T^2 de Hotelling proposée par Mason, Young et Tracy est qu'elle conserve une démarche unitaire avec la carte de contrôle T^2 . La détection est réalisée sur le carte de contrôle T^2 et ultérieurement, sans changer la statistique de contrôle, les informations nécessaires à l'interprétation sont fournies par la méthode de décomposition MYT.

Un inconvénient de la méthode est le grand nombre de facteurs $(p \times 2^{(p-1)})$ qu'on doit calculer dès que le nombre de variables devient important (par exemple pour une analyse avec 5 variables on doit calculer 80 facteurs et pour 10 variables 5120 facteurs). pour remédier à cet inconvénient Mason, Young et Tracy [80] ont proposé un algorithme en 5 étapes qui permet de réduire le nombre de calculs d'une façon significative.

Ultérieurement, Mason et Young [82] reviennent sur la décomposition MYT et proposent des méthodes qui peuvent améliorer la sensibilité de la statistique T^2 . Ils précisent l'importance de bien choisir un modèle de spécifications pour le processus dans la détection des sauts de la moyenne. Pour choisir le bon modèle, ils conseillent de demander l'avis des experts ou d'explorer attentivement les données avant de commencer l'analyse. Afin de bien détecter les dérives lentes de la moyenne, il est utile d'identifier les tendances sur les diagrammes des résidus standardisés obtenus à la suite d'une régression linéaire associée avec chaque facteur conditionné de la décomposition T^2 . Cette méthode est similaire à celle utilisée dans le diagramme sélection de la cause de Zhang (voir le paragraphe 2.3.6).

2.3.4 Algorithme BSA (Backward Selection Algorithm) et méthode de l'hyperplan

Une méthodologie de contrôle, en trois étapes, a été proposée par Chua et Montgomery [14]. Elle a été conçue pour détecter des échantillons hors contrôle, déterminer un sous-ensemble des variables hors contrôle et identifier finalement la variable ou les variables étant responsable(s) du dérèglement. Cette méthodologie représente une amélioration de la méthode de contrôle multivariée proposée par Chua [15] car elle prend en compte la corrélation entre les variables.

Les trois étapes de la méthode schématisée sur la figure II.4 sont:

- 1) utiliser la carte de contrôle MEWMA afin de détecter des dérèglements dans le cas multivarié;
- 2) si une situation hors contrôle à l'étape précédente est détectée, utiliser l'algorithme BSA (**B**ackward **S**election **A**lgorithm) afin d'identifier un groupe de variables hors contrôle;
- 3) utiliser la méthode de l'hyperplan afin d'analyser les variables.

Les observations multivariées sont reportées sur la carte de contrôle MEWMA. Si le processus est sous contrôle on continue les prélèvements. Dans le cas contraire, on vérifie le nombre des variables du processus. Si le nombre des variables est inférieur à cinq, l'algorithme BSA est utilisé pour réduire encore le nombre des variables en gardant uniquement les variables probablement responsables du dérèglement. Ces variables seront ultérieurement analysées par la méthode de l'hyperplan.

Le cas où le nombre de variables analysées est supérieur à cinq, l'algorithme BSA n'est plus utilisable et on peut passer directement à l'analyse par la méthode de l'hyperplan. Selon les conclusions tirées après l'analyse par la méthode de l'hyperplan les actions correctives nécessaires sont prises. Ensuite, un nouvel échantillon est prélevé et le cycle recommence.

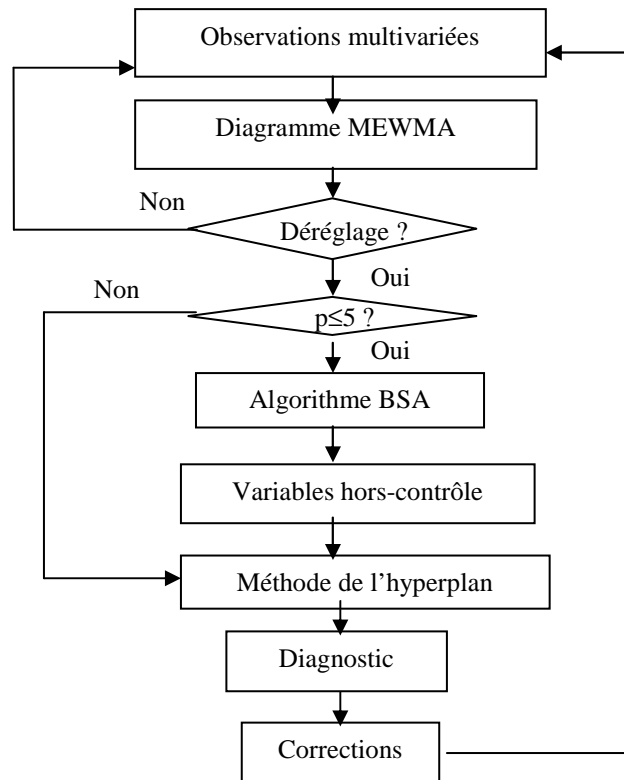


figure II.4 - Schéma de contrôle proposé par Chua et Montgomery

L'algorithme BSA a été conçu pour répondre à deux critères très importants pour tout algorithme de sélection efficace:

1. Le nombre de tests doit être le plus réduit possible
2. L'algorithme doit être conservatoire

Le premier critère, lié à la vitesse de calcul, assure que le temps de réponse peut être minimisé quand le nombre de tests est réduit.

Le deuxième critère garantit que le groupe des variables retenu pour l'analyse ne contient que des variables hors contrôle.

L'algorithme BSA divise les p variables en deux groupes de taille p_1 et p_2 , où $p_1 + p_2 = p$. La matrice de variance-covariance est divisée aussi en deux matrices, une avec p_1 variables et la deuxième avec p_2 variables. Pour l'ensemble de p variables, la statistique T^2 de Hotelling est:

$$T_p^2 = (\mathbf{X} - \bar{\mathbf{X}})^T \mathbf{S}^{-1} (\mathbf{X} - \bar{\mathbf{X}}) \quad (2.8)$$

où \mathbf{X} est le vecteur d'observations, $\bar{\mathbf{X}}$ est le vecteur de moyennes et \mathbf{S}^{-1} est l'inverse de la matrice de variance-covariance.

D'une manière similaire, pour le sous-groupe de p_1 variables, la statistique T^2 de Hotelling prend la forme suivante:

$$T_{p_1}^2 = (\mathbf{X}_{p_1} - \bar{\mathbf{X}}_{p_1})^T \mathbf{S}_{p_1}^{-1} (\mathbf{X}_{p_1} - \bar{\mathbf{X}}_{p_1}) \quad (2.9)$$

où X_{p_1} est le vecteur d'observations pour le sous-groupe de p_1 variables, \bar{X}_{p_1} est le vecteur des moyennes pour le sous-groupe de p_1 variables et $S_{p_1}^{-1}$ est l'inverse de la matrice de variance-covariance pour le sous-groupe de p_1 variables. On calcule la différence \mathbf{D} entre ces deux valeurs:

$$\mathbf{D} = T_p^2 - T_{p_1}^2 \quad (2.10)$$

Cette différence \mathbf{D} suit une loi de distribution de type khi-deux avec p_2 degrés de liberté. Par conséquent, on peut savoir si \mathbf{D} est statistiquement significative en la comparant avec le quantile d'ordre α de cette distribution. L'algorithme de sélection est basé, donc, sur la valeur de \mathbf{D} .

Si \mathbf{D} est statistiquement significative (selon le test de khi-deux), l'hypothèse selon laquelle le sous-groupe de p_1 variables est responsable de la situation hors contrôle est rejetée. Cela peut s'expliquer par le fait que ce groupe de variables n'a pas de contribution significative à la statistique T^2 . Pour le cas contraire, l'hypothèse doit être acceptée et on peut dire que le sous-groupe en cause contient les variables hors contrôle. Les opérations effectuées par l'algorithme BSA sont les suivantes:

- Etape 1: Calculer la statistique T_p^2 pour l'ensemble de p variables. Si une situation hors contrôle a été détectée passer à l'étape 2.
- Etape 2: Calculer, en éliminant une variable à la fois, toutes les statistiques de type T_{p-1}^2 . Calculer ensuite les différences $D_{p-1} = T_p^2 - T_{p-1}^2$. Retenir les variables pour lesquelles D_{p-1} est statistiquement non-significative. Noter $D_{\min(p-1)}$ la plus petite valeur D_{p-1} .
- Etape 3: Calculer, en éliminant deux variables à la fois, toutes les statistiques de type T_{p-2}^2 et les différences $D_{p-2} = T_p^2 - T_{p-2}^2$. Retenir les variables pour lesquelles D_{p-2} est statistiquement non-significative. Noter $D_{\min(p-2)}$ la plus petite valeur D_{p-2} .
- ⋮
- Etape p : Calculer toutes les statistiques de type T_1^2 et les différences $D_1 = T_p^2 - T_1^2$. Retenir les variables pour lesquelles D_1 est statistiquement non-significative. Noter $D_{\min(1)}$ la plus petite valeur D_1 .
- Etape $p+1$: Si toutes les distances D calculées à partir de l'étape 2 et jusqu'à l'étape p sont statistiquement significatives alors toutes les variables sont hors contrôle.
Si $D_{\min(i)} < D_{\min(j)}$, où $i < j$, ($i=1, \dots, p-1$ et $j=i+1, \dots, p$) et les variables du groupe i constituent un sous-groupe des variables du groupe j , alors le groupe des variables i est le groupe des variables hors contrôle. Dans le cas contraire, on incrémente i et on vérifie de nouveau cette condition.

En utilisant comme critère la réunion des groupes de variables hors contrôle, l'algorithme BSA comporte au maximum $\sum_{j=1}^{p-1} C_p^j$ tests. Sachant que la taille du groupe des variables après chaque étape peut être inférieure à p , le nombre de tests, en général, est inférieur au nombre de tests maximal.

Il est bien connu que l'ellipse de contrôle est un outil de diagnostic très utile pour le cas bivarié. Chua et Montgomery ont conçu dans le cas multivarié une méthode de diagnostic similaire à la méthode de l'ellipse de contrôle. Cette méthode porte le nom de méthode de l'hyperplan.

L'équation de l'hyper ellipsoïde de contrôle dans le cas multivarié est:

$$\mathbf{X}^T \mathbf{S}^{-1} \mathbf{X} = T_{\text{ref}} \quad (2.11)$$

où T_{ref} est la valeur de la limite de contrôle sur la carte T^2 de Hotelling;

Si pour une certaine observation la partie gauche de l'équation (2.11) est inférieure à T_{ref} , alors cette observation sera située à l'intérieur de l'hyper-ellipsoïde de contrôle et donc, elle sera sous contrôle. Si, par contre, la partie gauche de l'équation (2.11) est supérieure à T_{ref} , alors cette observation sera située à l'extérieur de l'hyper-ellipsoïde de contrôle et donc, elle sera hors contrôle. Dans ce cas, par l'intersection de l'hyper-ellipsoïde de contrôle avec des hyperplans qui passent par le point hors contrôle et parallèles à certains axes de coordonnées, on obtient des régions de contrôle elliptiques (figure II.5). Pour un nombre de p variables on aura un nombre de $p(p-1)/2$ cartes de contrôle elliptiques.

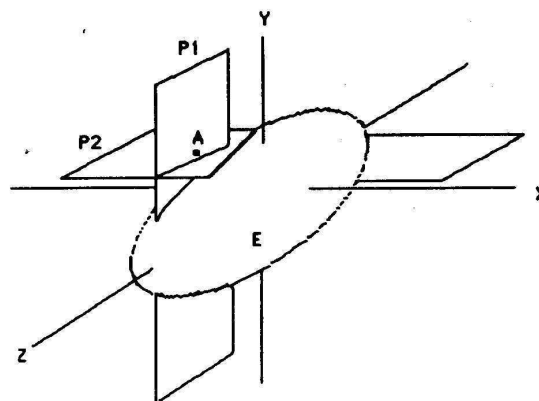


figure II.5 - Méthode de l'hyperplan

L'équation vectorielle d'un hyperplan est:

$$\mathbf{d}(\mathbf{x}-\mathbf{b}) = 0 \quad (2.12)$$

où \mathbf{d} est le vecteur directeur de l'hyperplan, \mathbf{x} est le vecteur d'observations et \mathbf{b} est une observation quelconque appartenant à l'hyperplan. Pour obtenir un hyperplan de direction parallèle avec un axe de coordonnées \mathbf{i} , le vecteur de direction \mathbf{d} doit être de la forme suivante:

$$\mathbf{d} = [0,0,\dots,i,\dots,0,0] \quad (2.13)$$

où $i = 1$ dénote l'axe de coordonnées envisagé.

Couper l'hyper-ellipsoïde de contrôle par l'hyperplan revient à résoudre un système de $p-1$ équations:

$$\begin{cases} \mathbf{X}^T \mathbf{S}^{-1} \mathbf{X} = T_{\text{ref}} \\ d(\mathbf{x} - \mathbf{b}) = 0 \end{cases} \quad (2.14)$$

Par des transformations algébriques le système de $p-1$ équations est réduit à un polynôme d'ordre 2. Ensuite, l'équation polynomiale est transformée en équation standard d'une ellipse. La méthode de l'hyperplan fournit $p(p-1)/2$ régions de contrôle elliptiques. Parfois, il se peut que l'équation elliptique soit dégénérée ou vide, ce qui signifie que le point hors contrôle est très éloigné de la région hyper-elliptique de contrôle. Dans ce cas, le nombre de cartes de contrôle sera plus réduit.

L'algorithme BSA semble très intéressant pour réduire le nombre de variables susceptibles d'avoir provoqué la situation hors contrôle. Malheureusement, son potentiel n'est mis en valeur que si le nombre de variables est inférieur ou égal à 5. Ceci est dû à l'erreur globale de première espèce qui augmente d'une manière significative dès qu'on a plus de 5 variables. Cet algorithme a été développé par d'autres auteurs sous différentes formes (voir la méthode de contributions de variables décrite au paragraphe 2.3.5).

La méthode de l'hyperplan nous semble pourtant assez difficile à mettre en œuvre, en raison d'une part du calcul laborieux nécessaire pour résoudre le système de p équations et pour transformer l'expression polynomiale de deuxième degré afin de retrouver l'équation elliptique, et d'autre part de la difficulté de l'interprétation des résultats. L'opérateur doit comparer et analyser un grand nombre des cartes de contrôles elliptiques (dans des conditions normales $p(p-1)/2$) et cela n'est pas très évident. En plus, selon les résultats obtenus par simulation, il est clair que la méthode de l'hyperplan ne donne pas de bons résultats dans le cas où les variables sont fortement corrélées car beaucoup de régions de contrôle elliptiques sont vides.

2.3.5 Méthode de calcul des contributions des variables

Runger, Alt et Montgomery [105] proposent trois méthodes différentes mais qui ont à la base la même métrique, pour détecter la source d'un dérèglement sur la carte de contrôle T^2 de Hotelling.

Leur première démarche est basée sur l'hypothèse que la contribution d'une variable à la statistique T^2 de Hotelling peut être estimée en calculant la valeur minimale de la statistique χ^2 (khi-deux) qui s'obtient après qu'on a modifié une seule variable dans le vecteur d'observations. Les variables qui nécessitent de grands modifications contribuent le plus à la statistique T^2 . Pour calculer la contribution de la variable x_i à la statistique T^2 il est nécessaire de déterminer la constante c_i qui minimise l'expression:

$$\left(\mathbf{X} - \frac{c_i \mathbf{e}_i}{\sqrt{\mathbf{e}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{e}_i}} \right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{X} - \frac{c_i \mathbf{e}_i}{\sqrt{\mathbf{e}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{e}_i}} \right) \quad (2.15)$$

où \mathbf{X} est le vecteur des observations, \mathbf{e}_i est le vecteur unitaire de l'axe de coordonnées i et $\boldsymbol{\Sigma}$ est la matrice de variance-covariance de \mathbf{X} ;

Dans l'équation (2.15) on considère que le vecteur de moyennes du processus est le vecteur nul. L'expression $e_i^T \Sigma^{-1} e_i$ représente un facteur d'échelle qui a été introduit pour que c_i soit interprétée comme une distance Euclidienne.

Notons: $Z = \Sigma^{-1/2} X$ et $v_i = \frac{\Sigma^{-1/2} e_i}{\sqrt{e_i^T \Sigma^{-1} e_i}}$. Alors le vecteur v_i est un vecteur unitaire dans

la direction de $\Sigma^{-1/2} e_i$ et l'équation (2.15) peut s'écrire:

$$(Z - c_i v_i)^T (Z - c_i v_i) \quad (2.16)$$

La constante c_i qui minimise (2.16) est la pente de la droite de régression de Z dans v_i , donc:

$$c_i = (v_i^T v_i)^{-1} v_i^T Z \quad (2.17)$$

mais comme $v_i^T v_i = 1$ on obtient:

$$c_i = v_i^T Z = \frac{e_i^T \Sigma^{-1} X}{\sqrt{e_i^T \Sigma^{-1} e_i}} \quad (2.18)$$

Ainsi, on retrouve la statistique utilisée par Hawkins, voir l'équation (2.1).

Pour déterminer la cause d'un dérèglement sur la carte T^2 , il faut comparer les amplitudes des coefficients $|c_i|$ (ou des coefficients c_i^2) pour $i=1\dots p$. Si $|c_i|$ est grand alors la variable x_i est la cause la plus probable du dérèglement. Du point de vue géométrique $|c_i|$ est la longueur de la projection orthogonale du vecteur Z dans le vecteur v_i . Ainsi:

$$|c_i| = \|P_i Z\| \quad (2.19)$$

où la matrice orthogonale de projection sur le vecteur v_i a été notée P_i . Les analogies géométriques sont facilitées après avoir effectué les transformations des variables dans Z et v_i car, ainsi la statistique T^2 devient simplement la distance Euclidienne au carré de Z par rapport au vecteur nul.

La deuxième démarche, qui n'est pas très différente de la précédente, consiste à calculer la distance:

$$D_i = T^2 - T_i^2$$

où T_i^2 est la statistique T^2 qu'on obtient après avoir remplacé x_i par c_i . Murphy a utilisé la même distance pour calculer la différence entre la statistique T^2 calculée à partir de toutes les variables et la statistique T^2 calculée à partir d'un sous-groupe de variables. Si D_i est grande alors x_i a une contribution importante dans le calcul de la statistique T^2 . Il est possible de démontrer sans difficulté que $D_i = c_i^2$, donc cette deuxième méthode n'est qu'une autre façon d'écrire et d'interpréter la méthode présentée antérieurement.

Considérons la variable standardisée $Z = \Sigma^{-1/2}X$. La troisième approche est basée sur un test d'hypothèse multivarié $H_0 : \mu = 0$ contre $H_1 : \mu \neq 0$. Si l'hypothèse nulle est vraie, cela implique que $v^T \mu = 0$ pour tous les vecteurs v . Il est connu que la statistique T^2 est le maximum parmi les statistiques utilisées pour tester l'hypothèse $H_0 : v^T \mu = 0$ contre l'hypothèse $H_1 : v^T \mu \neq 0$, c'est-à-dire:

$$T^2 = \max_v \left[\frac{(v^T Z)^2}{v^T v} \right]$$

La valeur maximale est obtenue pour tout multiple du vecteur $v = Z$. Une variable x_i est importante si le vecteur qui maximise T^2 est substantiellement dans la direction de e_i (après avoir transformé les vecteurs). Par conséquent, après avoir redimensionné les vecteurs pour avoir la norme unitaire, la mesure de la contribution de x_i dans la statistique T^2 est:

$$\frac{Z^T (\Sigma^{-1/2} e_i)}{\|Z\| \|\Sigma^{-1/2} e_i\|} = \frac{e_i^T \Sigma^{-1} X}{\sqrt{X^T \Sigma^{-1} X} \sqrt{e_i^T \Sigma^{-1} e_i}} \quad (2.20)$$

Puisque $X^T \Sigma^{-1} X$ dans l'équation (2.20) est une constante pour tout i , cette mesure de la contribution est équivalente avec celle obtenue précédemment.

Pour conclure, l'algorithme de détection proposé par Runger, Alt et Montgomery est très simple et repose essentiellement sur le calcul de la distance $D_i = T^2 - T_{(i)}^2$, où $T_{(i)}^2$ est la statistique T^2 obtenue après avoir éliminé la variable i dans le vecteur X (figure II.6).

Runger a démontré que $T_{(i)}^2 = T_i^2$, où T_i^2 est la statistique T^2 qu'on obtient après avoir remplacé la variable i du vecteur X avec la constante c_i . Il est recommandé de calculer pour chaque variable la métrique D_i et de retenir seulement les valeurs qui sont importantes pour l'analyse. Même si parfois une simple comparaison visuelle des valeurs pour D_i peut nous aider à trouver la variable responsable du dérèglement, l'utilisation de la statistique khi-deux peut nous donner aussi une estimation numérique pour le seuil à partir duquel cette valeur sera considérée statistiquement importante. L'hypothèse nulle selon laquelle x_i ne participe pas au dérèglement peut s'exprimer comme: $E(c_i) = E(v_i^T Z) = e_i^T \Sigma^{-1} \mu = 0$. Selon cette hypothèse il est facile de déduire, voir l'équation (2.18) que D_i a une distribution khi-deux avec un degré de liberté. Par conséquent, le quantile de 99% de cette distribution pourra être utilisé comme limite de contrôle. Si la vraie valeur de la matrice de variance-covariance Σ n'est pas connue, et si on l'estime à partir des données disponibles, alors la distribution F de Fisher avec 1 et $n-1$ degrés de liberté doit être utilisée pour le calcul de la limite de contrôle pour D_i . A la suite de l'utilisation de la matrice de variance-covariance, la métrique D_i diffère d'une façon importante du test t utilisé par Doganaksoy, Faltin et Tucker. Cette distance est équivalente à un test t seulement quand les variables sont indépendantes. Par rapport à la méthode de décomposition MYT le nombre de calculs est plus réduit mais on n'a pas la possibilité de détecter un changement survenu dans la relation entre les variables.

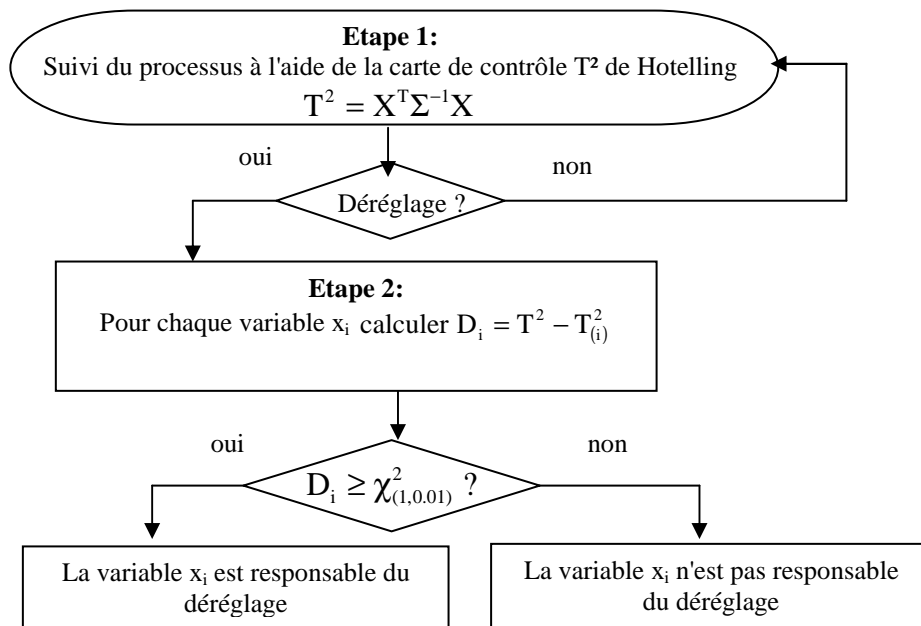


figure II.6 - Schéma de contrôle proposé par Runger, Alt et Montgomery

2.3.6 Carte de contrôle sélection de la cause

En 1984 Zhang [134] introduisait la carte de contrôle sélection de la cause. Cette carte de contrôle ressemble à la carte de régression de Mandel et son principe de construction n'est pas très éloigné de celui proposé par Hawkins que nous avons analysé dans le paragraphe 2.3.1. L'approche de Zhang est destinée aux processus en cascade ou aux processus multi-étapes. Comme la majorité des processus industriels sont très complexes, le processus étant la somme de plusieurs étapes ou sous-étapes, la méthodologie de Zhang est très pratique. A la base de la carte sélection de la cause sont les concepts de *Qualité Globale* et *Qualité Spécifique*. La *Qualité Globale* est la qualité qui résulte de l'étape courante et de toutes les étapes antérieures. La *Qualité Spécifique* est la qualité qui est obtenue à un moment donné. Evidemment la qualité globale est, en final, celle qui est perçue par le client. La carte de contrôle sélection de la cause a été conçue pour distinguer entre les causes assignables contrôlables et les causes assignables incontrôlables. Les causes assignables contrôlables sont celles qui affectent l'étape courante du processus mais n'affectent pas les étapes précédentes. Les causes assignables incontrôlables sont les causes qui affectent les étapes antérieures et ne peuvent pas être contrôlées dans l'étape courante.

Pour simplifier prenons le cas d'un processus en cascade ayant deux étapes. Soit X la variable observée à l'étape 1 et Y la variable observée à l'étape 2. Dans un premier temps on identifie la relation existante entre Y et X soit par des méthodes empiriques, soit par des méthodes théoriques. Le modèle reliant les deux variables peut prendre différentes formes. Un des modèles les plus utiles est la régression linéaire simple:

$$Y_i = b_0 + b_1 X_i + \varepsilon_i \quad (2.21)$$

où les valeurs ε_i sont considérées indépendantes et identiquement distribuées selon une distribution normale de moyenne nulle et écart-type égal à 1. Le modèle nécessaire pour la construction de la carte de contrôle sélection de la cause ne doit pas être

obligatoirement linéaire. La technique sélection de la cause peut s'appliquer aussi aux modèles non-linéaires pourvu que les hypothèses relatives aux termes d'erreur soient respectées. Zhang affirme que si le processus est sous contrôle (c'est-à-dire il n'y a pas des causes assignables) les valeurs Y_i sont indépendantes et distribuées selon une distribution normale de moyenne μ_i et écart-type σ_i , où $\mu_i = f(X_i)$ et $\sigma_i = g(X_i)$. Si les fonctions f et g sont connues, alors par la transformation:

$$Z_i = \frac{Y_i - \mu_i}{\sigma_i} \quad (2.22)$$

on obtient des valeurs Z_i qui sont indépendantes et distribuées selon une distribution normale de moyenne nulle et écart-type égal à 1.

Zhang appelle les valeurs Z_i *les valeurs sélection de la cause* parce qu'elles sont les valeurs de Y_i ajustées pour les effets de X_i . Les limites de contrôle sur la carte sont fixées à ± 3 .

En pratique, la relation entre Y et X n'est pas connue et les valeurs μ_i et σ_i doivent être estimées à partir d'un échantillon de taille n . L'estimateur de Y_i est noté \hat{Y}_i . La carte sélection de la cause est, en effet, une carte classique (*Shewhart, EWMA ou CUSUM*) appliquée aux résidus ($Z_i = Y_i - \hat{Y}_i$). L'échantillon initial de n données est utilisé pour estimer la relation entre Y et X et pour le calcul des limites de contrôle. La valeur centrale sur la carte sélection de la cause est calculée avec la formule:

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i) \quad (2.23)$$

On peut déduire facilement que $\bar{Z} = 0$ si la méthode des moindres carrés est utilisée pour la régression de Y en X .

Pour calculer les limites de contrôle sur la carte, Zhang a estimé la variance à partir de la moyenne mobile d'ordre 1 appliquée aux résidus.

$$\begin{aligned} \text{LCS} &= \bar{Z} + 3\hat{\sigma} = \bar{Z} + A_3 \overline{\text{MR}} \\ \text{LCI} &= \bar{Z} - 3\hat{\sigma} = \bar{Z} - A_3 \overline{\text{MR}} \end{aligned} \quad (2.24)$$

$$\text{où } \overline{\text{MR}} = \frac{1}{n-1} \sum_{i=1}^{n-1} \text{MR}_i ; \text{MR}_i = |Z_{i+1} - Z_i| \text{ et } A_3 = 2.66.$$

La carte sélection de la cause s'utilise conjointement avec les cartes classiques pour le diagnostic du processus. Les règles de décision pour le cas de deux variables sont présentées dans le tableau II.1.

Un des avantages de la carte sélection de la cause est qu'elle ne nécessite pas une distribution multivariée normale pour les variables, ce qui nous permet d'analyser aussi des variables liées par des relations non-linéaires. Elle peut aussi localiser la source d'un dérèglement ce qui la rend plus performante que la carte T^2 Hotelling. L'inconvénient de la méthode est le grand nombre de cartes qu'on doit observer en même temps quand le nombre de variables est important.

tableau II.1

Carte pour X	Carte pour Z	Interprétation
déréglage	déréglage	les deux variables sont hors contrôle
déréglage	pas de déréglage	la variable X est hors contrôle
pas de déréglage	déréglage	la variable Y est hors contrôle
pas de déréglage	pas de déréglage	les deux variables sont sous contrôle

2.3.7 Carte de contrôle multivariée des profils

Fuchs et Benjamini [33] ont proposé une méthodologie de contrôle multivariée pour la détection de la source d'une situation hors contrôle. Ils ont voulu conserver la simplicité de la carte de contrôle de type Shewhart et ont fait figurer sur une seule carte de contrôle deux caractéristiques:

- ✓ une caractéristique univariée (la variable centrée et réduite):

$$d_{ij} = \frac{\bar{x}_{ij} - \mu_i}{\sigma_i} \quad (2.25)$$

où $\bar{x}_{ij} = \frac{1}{n} \sum_{l=1}^n x_{ijl}$ est la moyenne de la variable i dans l'échantillon j de n observations, x_{ijl} est l'observation l ($l=1 \dots n$) de la variable i dans l'échantillon j , μ_i est la moyenne de la variable i et σ_i est l'écart-type de la variable i .

- ✓ une caractéristique multivariée (la statistique T^2 de Hotelling):

$$T_j^2 = (\bar{x}_j - \mu)^T \Sigma^{-1} (\bar{x}_j - \mu) \quad (2.26)$$

où $\bar{x}_j = (\bar{x}_{1j}, \dots, \bar{x}_{pj})^T$ est le vecteur des moyennes de l'échantillon j , $\mu = (\mu_1, \dots, \mu_p)^T$ est le vecteur de moyennes et Σ est la matrice de variance-covariance, avec $\Sigma_{ii} = \sigma_i^2$ dans l'équation (2.25)

Fuchs et Benjamini ont appelé cette nouvelle carte de contrôle "carte de contrôle multivariée de profils". L'interprétation visuelle de la carte de contrôle est très intuitive. L'information relative à une certaine variable est schématisée à l'aide d'un symbole (un rectangle vertical) et l'information concernant le groupe des variables est représentée par une ligne horizontale qui constitue la base de ces rectangles (figure II.7).

L'amplitude et l'orientation du symbole rectangulaire nous informe sur l'ordre de grandeur et le sens de l'écart entre la moyenne d'une variable et sa valeur de référence. Après avoir établie la position de la ligne horizontale (information multivariée) sur la carte, des rectangles pour chaque variable sont positionnés par rapport à cette ligne. Le rectangle sera positionné au-dessus de la ligne horizontale si l'écart est positif ou en-dessous de cette ligne si l'écart est négatif. La hauteur du rectangle est en concordance avec l'amplitude de l'écart d_{ij} , voir l'équation (2.25), mais ne dépasse pas la valeur de 4. Si l'écart normalisé dépasse 2, le rectangle sera hachuré. S'il dépasse 3, le rectangle sera coloré en noir. Tout comme pour la carte de Shewhart les symboles sont placés d'une

façon séquentielle sur l'axe des abscisses (temps). Le positionnement sur l'axe vertical est donné par la statistique T^2 de Hotelling. Deux valeurs sont représentées sur cet axe: $T^2_{0,997}$ et $T^2_{0,95}$ le quantile 99,7% et respectivement 95% de la distribution de T^2 .

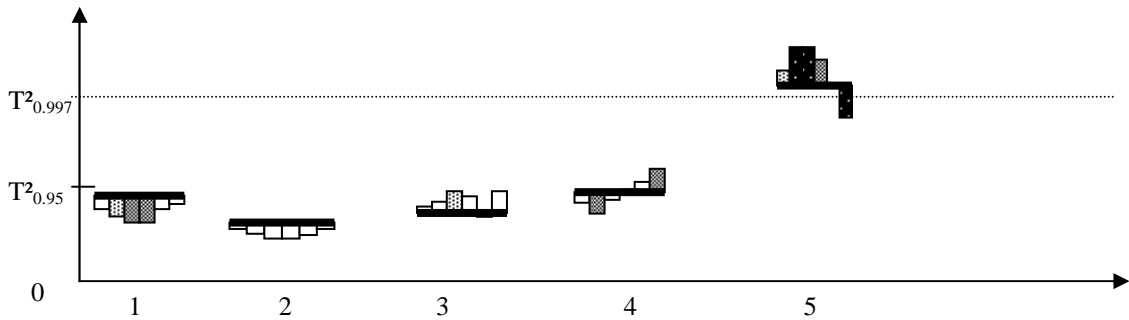


figure II.7 - la carte de contrôle multivariée des profils

Quand le processus est sous contrôle, seule une ligne horizontale sera visible en bas de la carte. Quand le processus est hors contrôle plusieurs informations d'alerte seront affichées sur la carte de contrôle: la hauteur du symbole rectangulaire augmente et la couleur du symbole passe au noir à travers plusieurs nuances de gris. Par conséquent, le principe de construction de la carte aide à identifier la source du dérèglement. Le fait de représenter sur la même carte de contrôle des variables individuelles et des informations de groupe nous donne aussi la possibilité de remarquer visuellement un changement de relation entre les variables.

Le dérèglement multivarié est mis en évidence sur la carte de contrôle multivariée de profils quand la statistique T^2 dépasse la limite critique $T^2_{0,997}$ ou quand deux observations successives dépassent la limite critique $T^2_{0,95}$. Une certaine caractéristique est considérée hors contrôle si pour deux observations successives l'écart normalisé dépasse la valeur de 3 (rectangle noir) et la valeur de 2 (rectangle hachuré). Une observation est hors contrôle également quand pour trois observations successives la valeur de l'écart normalisé est supérieure à 2. Ces règles de décision pour la statistique T^2 sont en concordance avec celles qui signalent des situations hors contrôle dans le cas univarié.

Comme nous l'avons remarqué dans le paragraphe 2.3.2 les indicateurs univariés ne donnent pas toujours un diagnostic fiable dans le cas multivarié car la corrélation entre les variables n'est pas prise en compte. L'avantage de la méthode proposée par Fuchs et Benjamini est de combiner ces indicateurs univariés avec des indicateurs multivariés et cela sous une forme graphique très intuitive pour l'interprétation (beaucoup plus facile à utiliser que le diagramme "polyplot" de Blazek et al.[6], ou le diagramme "starplot" du logiciel STATGRAPHICS 3.0 [115]). Des règles de décision sont proposées afin de déterminer la variable probablement responsable d'un dérèglement détecté sur la carte de contrôle. Fuchs et Benjamini montrent aussi que la carte de contrôle multivariée des profils peut être adaptée pour le suivi de la variabilité du processus. L'inconvénient de la méthode est que dans le cas où le nombre de variables est important l'efficacité du diagnostic est diminuée. En plus, la représentation graphique devient difficile à gérer dans ce cas.

2.3.8 Les méthodes de projection

Quand les données sont fortement corrélées, l'utilisation des méthodes basées sur l'inverse de la matrice de variance-covariance n'est plus possible car la matrice de variance-covariance est singulière. Une solution possible à ce problème a été proposée par Houshmand et Javahery [54]. Elle repose, voir l'annexe A2, sur le calcul des coefficients de régression biaisés.

Les chercheurs ont vu dans l'ACP (l'Analyse en Composantes Principales) un moyen de sortir de cet impasse. L'utilisation de l'ACP pour le suivi et l'analyse des processus multivariés dans un espace de dimension réduite est une technique qui donne des bons résultats [55, 65, 67, 129] et présente de nombreuses applications pratiques [64, 66, 88]. Si on dispose d'un ensemble de données caractérisant le fonctionnement normal du processus (quand seulement les causes communes de variation sont présentes), on démontre qu'un nombre réduit de composantes principales expliquent presque toute la variabilité du processus.

Les composantes principales (ou axes factoriels) représentent un ensemble de variables artificielles (le résultat d'une combinaison linéaire des variables originales) non corrélées entre elles et orthogonales deux à deux. Les coordonnées des observations originales (regroupées dans une matrice X de dimension $n \times p$) sur l'axe factoriel a sont données par l'équation suivante:

$$t_a = p_a^T (X - \mu) = \sum_{j=1}^n p_{a,j} (X_j - \mu_j) \quad (2.27)$$

où p_a est le vecteur propre de la matrice S (de variance-covariance estimée à partir des données disponibles X) associé à la valeur propre λ_a , $p_{a,j}$ est la j -ème composante du vecteur p_a , X_j est la valeur de la variable j dans le vecteur d'observations X et μ_j est la moyenne de la variable j .

Par conséquent, une approximation de la matrice de données initiales X à l'aide de A composantes principales est donnée par l'équation ci-dessous:

$$X = t_1 p_1^T + t_2 p_2^T + \dots + t_A p_A^T + E = T_A P_A^T + E = \hat{X} + E \quad (2.28)$$

où E est appelée matrice résiduelle et \hat{X} est l'estimation de la matrice X à l'aide de A composantes principales. Si le nombre des composantes principales A est bien choisi il ne reste pas beaucoup d'information utile dans la matrice E (E représente dans ce cas une erreur aléatoire). Plusieurs méthodes existent pour déterminer le nombre des composantes principales significatives A . Une solution consiste à ajouter des composantes principales jusqu'à ce qu'un pourcentage de la variation totale est expliquée. Cette méthode généralement donne un nombre de composantes principales surestimé. Une meilleure solution a été proposée par Wold [130] et consiste à soustraire un certain nombre d'observations de l'ensemble de données, construire un modèle ACP sur les données restantes et ensuite évaluer la somme carré des erreurs de prédiction du modèle pour les données soustraites. Cette procédure est réitérée jusqu'à ce que chaque observation a été éliminé au moins une fois. Le nombre optimal de composantes principales A est celui qui minimise la somme carré des erreurs de prédiction.

Une fois qu'on a établi le modèle ACP, les nouvelles observations x_{nouveau} sont projetées sur l'hyperplan de dimension A défini par les vecteurs propres p_a ($a=1, \dots, A$) et leurs

coordonnées ($t_{A,\text{nouveau}} = P_A^T x_{\text{nouveau}}$) et résidus ($e_{\text{nouveau}} = x_{\text{nouveau}} - \hat{x}_{\text{nouveau}}$) sont calculés. Des cartes de contrôle multivariées, comme la carte T^2 de Hotelling par exemple, peuvent être utilisées pour le suivi dans l'espace des composantes principales. La formule d'équivalence entre la statistique T^2 et les composantes principales est:

$$T^2 = \sum_{a=1}^A \frac{t_a^2}{\lambda_a} = \sum_{a=1}^A \frac{t_a^2}{s_a^2} \quad (2.29)$$

Donc la statistique T^2 pour les variables originales est équivalente à une statistique T^2 pour les composantes principales normalisées. La limite supérieure de contrôle pour cette carte est calculée à l'aide de la distribution F de Fisher [123]. Cette carte de contrôle détecte seulement les causes assignables qui se manifestent dans le plan de A composantes principales retenues dans le modèle. Si la carte T^2 signale l'apparition d'une cause assignable c'est parce-que la distance entre une certaine observation et le centre de gravité des points sous contrôle est trop grande. La validité du modèle ACP du processus reste n'est pas mise en cause dans ce cas.

A l'apparition d'une nouvelle cause assignable (qui n'a pas été présente dans les données contribuant à la construction du modèle PCA) les observations correspondantes vont s'écarter de l'hyperplan de dimension A . Ce type de situation peut être détectée en calculant la somme carrée des erreurs de prédiction (SCEP), qui est nommée aussi statistique Q :

$$Q = \sum_{i=1}^m (x_{\text{nouveau},i} - \hat{x}_{\text{nouveau},i})^2 = (x - \hat{x})^T (x - \hat{x}) \quad (2.30)$$

Quand le processus est sous contrôle, la statistique Q représente des fluctuations aléatoires qui ne peuvent pas être expliquées par le modèle ACP. Quand une nouvelle cause assignable de variation se manifeste (par un changement dans le vecteur des moyennes ou dans la matrice de variance-covariance du processus) elle sera mise en évidence par une valeur élevée de la statistique Q . La limite de contrôle supérieure pour cette statistique est calculée à partir de l'ensemble de données de référence, Jackson [56, 89].

Très souvent, en pratique, on veut inclure un deuxième ensemble de variables (Y) dans la procédure de suivi. Ces variables peuvent être, par exemple, des caractéristiques de qualité du produit où des variables mesurant la productivité. Malheureusement, pour des considérations d'ordre pratique, ce type de variables peuvent être enregistrées moins fréquemment que les variables de processus. Par conséquent, il est souhaitable d'utiliser l'information contenue dans les variables de processus (X) pour le suivi et la détection des changements dans les variables de la sortie (Y). La méthode de Projection dans les Structures Latentes (PSL), ou PLS (Projection to Latent Structures) en anglais, semble être la meilleure solution pour ce type de problèmes. La méthode PSL (qui ressemble beaucoup avec la méthode ACP) réduit simultanément la dimension des espaces X et Y en cherchant les vecteurs propres des deux espaces qui sont les plus corrélés. En effet, l'hyperplan de dimension réduite trouvé à l'aide de la méthode PSL dans l'espace X est celui qui est le plus prédictif pour l'espace Y . Les vecteurs propres pour la méthode PSL sont trouvés à l'aide d'une méthode itérative et on peut démontrer qu'ils correspondent aux valeurs propres de la matrice $X^T Y Y^T X$, où $Y^T X$ est la matrice de covariance entre X et Y (Hoskuldsson [52]). A part ces aspects théoriques, le méthode de suivi du

processus en utilisant la projection dans les structures latente ne diffère guère de la méthode de suivi basée sur l'analyse en composante principale.

Détecter la source d'un dérèglement dans l'espace des composantes principales reste pourtant difficile, car les nouvelles variables sont des combinaisons linéaires des variables originales. Ainsi même si on sait qu'une certaine composante principale est responsable du dérèglement on ne peut pas savoir quelle est la variable ou quelles sont les variables responsables, sauf pour le cas où, voir l'annexe A3, les composantes principales peuvent s'interpréter comme des grandeurs physiques du processus.

Kurti et MacGregor [65] ont proposé une démarche originale pour le diagnostic d'un processus hors contrôle. Si à l'instant \mathbf{k} la carte de contrôle T^2 signale une situation hors contrôle, on calcule les variables centrées et réduites avec la formule :

$$\frac{(x_{k,j} - \mu_j)}{s_j} \quad (2.31)$$

où $x_{k,j}$ est l'observation \mathbf{k} de la variable \mathbf{j} , μ_j est la moyenne de la variable \mathbf{j} et s_j est l'écart-type de la variable \mathbf{j} . Les variables centrées et réduites sont ensuite représentées sur un diagramme et celles ayant des valeurs supérieures à ± 3 sont la cause la plus probable du dérèglement. Cette méthode, comme on l'a déjà constaté, ne donne pas de bons résultats si les variables sont fortement corrélées. Pour remédier à cet inconvénient Kurti et MacGregor proposent l'utilisation de cartes pour les composantes principales normalisées (t_a/s_a) car les composantes principales sont indépendantes. On retient les composantes principales normalisées importantes (les limites de type Bonferroni peuvent être utilisées). Ensuite, on détermine la contribution des variables au calcul des composantes et on retient les variables ayant de grandes contributions comme étant les plus probablement responsables du dérèglement. La contribution de la variable X_j à la composante principale \mathbf{a} est :

$$p_{a,j}(X_j - \mu_j) \quad (2.32)$$

Les contributions importantes et ayant le même signe que les composantes principales normalisées seront retenues pour identifier la variable qui a subi des modifications durant l'expérimentation.

Très souvent, en pratique, plusieurs composantes principales normalisées ont des valeurs importantes. Il sera alors utile de déterminer la contribution totale des variables au calcul des composantes principales normalisées. Si on a K composantes principales normalisées importantes, pour chaque variable X_j on calcule sa contribution au calcul de la composante principale normalisée \mathbf{a} avec la formule :

$$\text{cont}_{a,j} = \frac{t_a}{s_a} p_{a,j}(X_j - \mu_j) = \frac{t_a}{\lambda_a} p_{a,j}(X_j - \mu_j) \quad (2.33)$$

et ensuite on calcule la contribution totale de la variable X_j avec la formule :

$$\text{CONT}_j = \sum_{a=1}^k (\text{cont}_{a,j}) \quad (2.34)$$

Cette méthode d'identification de la variable ou des variables responsable(s) d'un dérèglement, basée sur le calcul des contributions, présente pourtant deux inconvénients majeurs: on ne peut pas quantifier ce qu'on comprend par "contribution importante" car on n'a pas une limite de contrôle clairement définie; on a des difficultés à déterminer d'une façon précise la cause d'un dérèglement dans le cas de la propagation des erreurs. Westerhuis et al. [126] donnent un exemple de ce type de propagation d'erreurs et proposent des limites de contrôle pour les diagrammes de contributions.

Les méthodes de projection ont trouvé beaucoup d'applications dans les domaines industriels, comme l'industrie chimique ou électronique (les processus "batch"), où un grand nombre de variables (qui peuvent être corrélées) sont contrôlées. Les résultats obtenus sont meilleurs par rapport à la carte T^2 Hotelling. Pourtant, la détection de la variable responsable d'un dérèglement s'avère une tâche assez compliquée car dans la plupart des cas des connaissances approfondies du processus sont nécessaires pour interpréter les résultats obtenus.

2.3.9 Méthodes de détection directionnelles

Dans ce paragraphe nous avons regroupé quelques méthodes d'identification de la variable responsable du dérèglement basées sur le décryptage de la "signature" caractérisant d'une manière unique chaque cause assignable. Par "signature" on doit comprendre le regroupement des points observations, caractérisant la présence d'une cause assignable, dans une certaine zone bien définie dans l'espace multidimensionnel.

Lafaye de Micheaux [68, 69] affirme que les points associés à une cause assignable s'organisent autour d'une direction privilégiée passant par le centre de gravité du nuage des points caractérisant le fonctionnement sous contrôle du processus (figure II.8). Lafaye de Micheaux [68, 69] propose une méthode qui consiste à identifier à quelle direction peut être associé un point hors contrôle (détecté par la carte de contrôle T^2 de Hotelling).

La proximité entre un point hors contrôle et une certaine direction peut être évaluée de deux façons:

- ✓ par l'abscisse du point projeté sur cette direction;
- ✓ par l'angle au centre θ que forme le point avec cette direction;

Le premier indice de proximité donne lieu à une "variable cause" qui peut être suivie à l'aide d'une carte de contrôle classique.

Le deuxième indice de proximité peut être exprimé en unités d'angle (degré, radian,...) ou par le cosinus de l'angle. Un inconvénient de cette démarche est que la distribution des valeurs de l'angle ou du cosinus de l'angle n'est pas gaussienne d'où une certaine difficulté pour établir les limites de contrôle sur la carte. Pour remédier à cet inconvénient Lafaye de Micheaux donne les valeurs des limites de contrôle pour un nombre de variables compris entre 2 et 100 et pour différents risques de première espèce α entre 0.0001 et 0.1. Un autre inconvénient lié à l'utilisation de l'angle θ (ou son cosinus) pour la détection des points hors contrôle est que l'efficacité de la détection est fonction de la dimension de l'espace des variables.

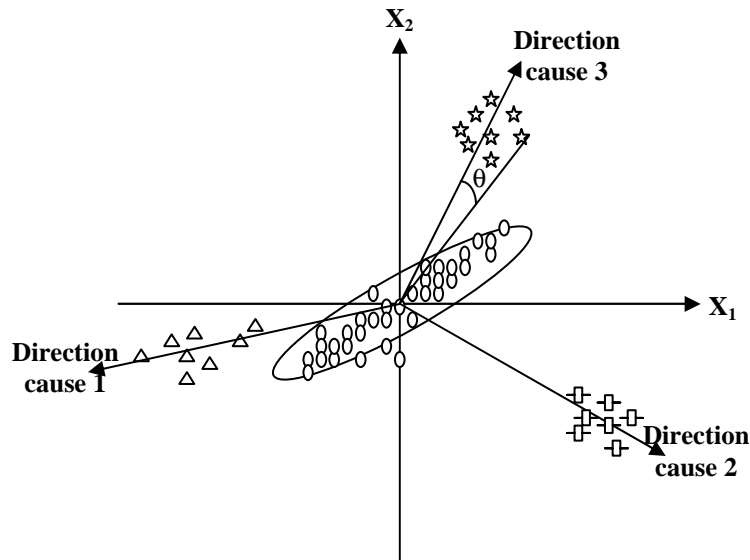


figure II.8 - Directions "spécifiques" dans l'espace multivarié

La mise en œuvre de la méthode commence par une analyse historique des données afin d'identifier les directions des causes assignables. Cette analyse comporte trois étapes:

- ✓ on établit d'abord une distinction entre les points sous contrôle et les points hors contrôle (pour cela on peut utiliser une procédure itérative basée sur l'élimination des points jugés hors contrôle par une carte T^2 de Hotelling);
- ✓ on regroupe ensuite par causes les observations déclarées hors contrôle (ce regroupement s'appuie à la fois sur la connaissance pratique du processus et sur une méthode de classification adaptée);
- ✓ on détermine enfin la direction "moyenne" à associer à chaque groupe défini à l'étape précédente.

Les directions propres aux causes assignables étant identifiées, on dispose des coefficients (abscisses de projection ou des cosinus d'angle) utilisables pour la détection. Dans la phase opérationnelle de contrôle, ces coefficients sont utilisés pour chaque nouvelle observation déclarée hors contrôle par la carte T^2 de Hotelling. L'identification de la cause assignable associée à un point hors contrôle se fait de la manière suivante:

- ✓ si une seule "variable cause" déclare le point hors contrôle alors la cause assignable est identifiée;
- ✓ si plusieurs variables cause sont candidates, on retient la (ou les) cause(s) associée(s) à la valeur d'angle la plus significative;
- ✓ si aucune variable cause n'est pas candidate, il s'agit sans doute d'une nouvelle cause qui n'a pas encore été répertoriée. L'identification se fera dans ce cas par tâtonnement. Cette observation permettra de définir une nouvelle direction de cause à intégrer au système de reconnaissance. Ainsi cette cause pourra être reconnue par la suite.

Parfois, néanmoins, il est difficile de distinguer ces directions dans le plan des variables originales. En plus, quand le nombre de variables est supérieur à trois il est impossible d'envisager une inspection visuelle. Pour remédier à cet inconvénient Lafaye de Micheaux [68] propose l'utilisation de l'espace des composantes principales.

Yoon et MacGregor [133] proposent une méthodologie de contrôle basée sur le changement, enregistré dans la présence d'une cause assignable, de l'orientation du vecteur d'observation dans l'espace ACP (l'espace du modèle et l'espace résiduel). La méthode peut s'utiliser conjointement avec les diagrammes de contribution des variables et aide à améliorer l'identification de la variable responsable du dérèglement. Yoon et MacGregor considèrent que les causes assignables sont de deux types: les *causes assignables simples* (leur effet se manifeste au niveau d'une seule variable et il ne se propage pas dans les autres variables) et les *causes assignables complexes* (sont causées par une certaine variable mais leur effet se propage et perturbe aussi les autres variables). L'identification de la variable responsable pour une cause assignable est relativement facile à l'aide des diagrammes de contributions. Par contre, l'identification de la variable responsable du dérèglement dans le cas des causes assignables complexes est plus difficile car sur le diagramme de contributions beaucoup de variables présentent de valeurs importantes. Pour remédier à cette situation, il est nécessaire d'avoir plus d'information sur la nature des causes assignables ou sur les relations entre les variables. Cette information est extraite d'une base de données où l'historique de l'évolution du processus se trouve enregistré.

Le principe de construction de la méthode de Yoon et MacGregor est très simple. Ils partent de l'hypothèse qu'une cause assignable complexe a une "signature" initiale (ou caractéristique lui étant propre et l'identifiant d'une manière unique), une trajectoire d'évolution temporelle et finalement une deuxième "signature" dans un état d'équilibre stable (figure II.9).

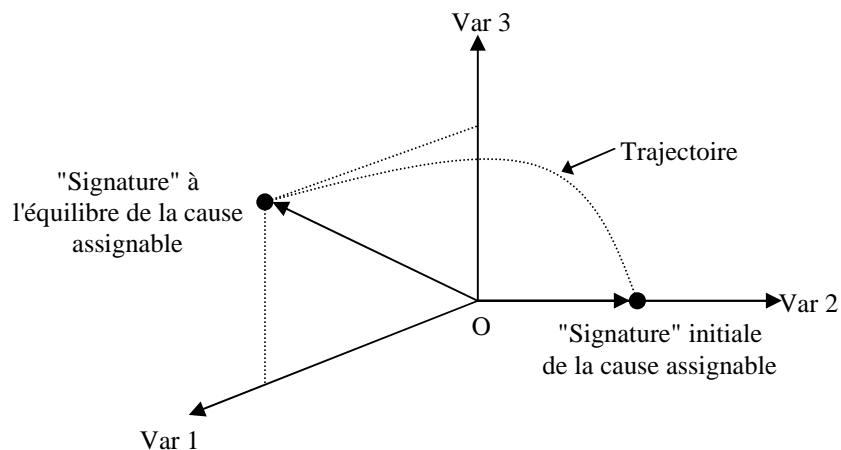


figure II.9 - Evolution de la "signature" de la cause assignable

La "signature" initiale de la cause assignable peut donner une indication précieuse au sujet de la variable responsable du dérèglement car elle n'est pas affectée par la propagation d'erreurs. Malheureusement, en pratique, il arrive qu'elle n'est pas détectée à temps et, par conséquent, la trajectoire de la transition et l'état final

d'équilibre sont utilisés comme une alternative pour caractériser les causes assignables complexes.

Si on note par x^n le vecteur d'observations juste avant l'apparition de la cause assignable, alors la présence d'une cause assignable i peut être modélisée par l'équation suivante: $x = x^n + f_i$ (où x est le nouveau vecteur d'observations et f_i est le vecteur de la cause assignable). Le vecteur de la cause assignable peut être décomposé en utilisant l'approche ACP en deux parties: le vecteur \hat{f}_i qui se trouve dans l'espace réduit défini par l'ACP et le vecteur \tilde{f}_i qui se trouve dans l'espace des résidus. Ces deux vecteurs sont, par la suite, normalisés pour qu'ils ne soient pas influencés par leurs amplitudes:

$$\hat{f}_i^0 = \frac{\hat{f}_i}{\|\hat{f}_i\|} \quad \tilde{f}_i^0 = \frac{\tilde{f}_i}{\|\tilde{f}_i\|} \quad , \text{ avec } \|\tilde{f}_i\| \neq 0 \text{ et } \|\hat{f}_i\| \neq 0 \quad (2.35)$$

Une base de données qui regroupe tous les vecteurs de "signature" pour tout $i = 1, 2, \dots, r$ est construite de la manière suivante:

$$\hat{F} = [\hat{f}_1^0, \hat{f}_2^0, \dots, \hat{f}_r^0] \quad \text{et} \quad \tilde{F} = [\tilde{f}_1^0, \tilde{f}_2^0, \dots, \tilde{f}_r^0] \quad (2.36)$$

Les deux matrices dans l'équation (2.36) contiennent toute l'information disponible concernant les r causes assignables dans les deux espaces (l'espace du modèle et l'espace résiduel).

Une fois qu'on a identifié ces deux matrices, la procédure de contrôle est très simple: on décompose tout nouveau vecteur d'observation x_{nouveau} en deux composantes \hat{x}_{nouveau} et $\tilde{x}_{\text{nouveau}}$ qui seront, par la suite, normalisées:

$$\hat{x}^0 = \frac{\hat{x}_{\text{nouveau}}}{\|\hat{x}_{\text{nouveau}}\|} \quad \text{et} \quad \tilde{x}^0 = \frac{\tilde{x}_{\text{nouveau}}}{\|\tilde{x}_{\text{nouveau}}\|} \quad (2.37)$$

et ensuite on mesure l'angle entre les vecteurs "signatures" de la base de données (les colonnes des matrices \hat{F} et \tilde{F}) et le nouveau vecteur d'observations (\hat{x}^0 et \tilde{x}^0) pour identifier la cause de l'anomalie. La figure II.10 illustre la relation existante entre le nouveau vecteur d'observations (x_{nouveau}) et le vecteur "signature" f_i correspondant à une cause assignable antérieurement détectée dans la variable x_1 .

Le cosinus de l'angle (θ) formé entre le nouveau vecteur d'observations et un parmi les vecteurs "signature" donne une mesure de la colinéarité existante entre eux.

Dans l'espace des résidus on a:

$$\cos \tilde{\theta}_1 = (\tilde{x}^0)^T \tilde{f}_1^0, \text{ où } -1 \leq (\tilde{x}^0)^T \tilde{f}_1^0 \leq 1 \quad (2.38)$$

Si la valeur de cosinus s'approche de 1 cela veut dire que le nouveau vecteur d'observations est presque colinéaire avec la direction de la cause assignable. Par conséquent, quand la carte T^2 ou la carte Q signalent une anomalie, la nature de la cause assignable peut être identifiée comme celle pour laquelle la valeur du cosinus est proche de 1, ou comme étant celle pour laquelle on a la valeur maximale du cosinus:

$$\max_i \left[(\tilde{x}^0)^T \tilde{F} \right] = \max_i \left[(\tilde{x}^0)^T \tilde{f}_1^0, (\tilde{x}^0)^T \tilde{f}_2^0, \dots, (\tilde{x}^0)^T \tilde{f}_r^0 \right] \quad (2.39)$$

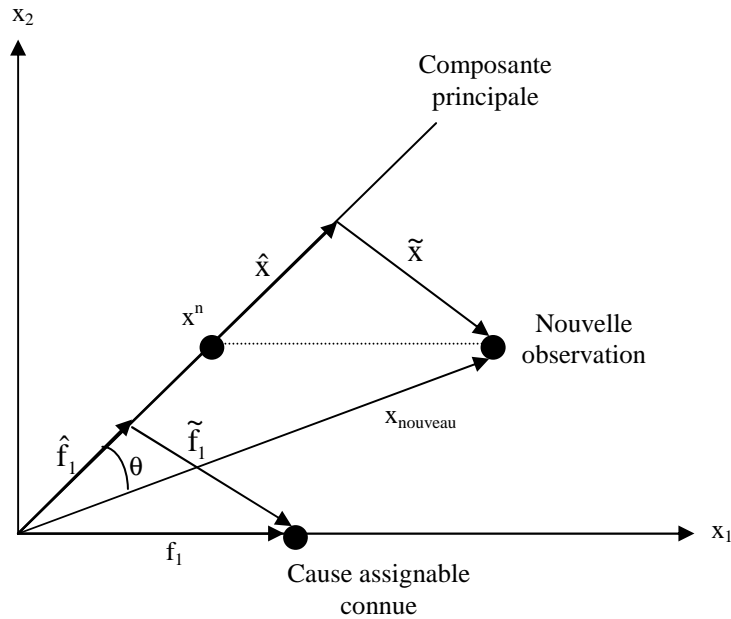


figure II.10 -Décomposition de la cause assignable et du nouveau vecteur d'observations

La composante cosinus du modèle ACP est calculée d'une manière similaire. L'identification de la cause assignable est faite par l'utilisation conjointe des diagrammes pour les valeurs cosinus dans l'espace de résidus et dans l'espace du modèle. L'avantage de la méthode est la détection non seulement des causes assignables simples mais aussi des causes assignables complexes. L'inconvénient de la méthode est que l'utilisation de la "signature" de la cause assignable dans un état d'équilibre induit inévitablement un certain retard entre le moment de l'apparition de la cause et l'identification.

2.4 Conclusion

Il est difficile de dire quelle est, parmi toutes ces techniques, la meilleure. Chacune a des avantages et des inconvénients. Certaines techniques, même si elle ne sont pas très performantes dans la localisation de la source d'un dérèglement, sont intéressantes parce qu'elles ont ouvert le chemin vers d'autres approches plus complexes et plus pratiques. On peut mentionner ici, à titre d'exemple, la méthode de régression multiple de Hawkins [46] et la méthode de calcul de la contribution des variables à la statistique T^2 Hotelling utilisant le test t (Student) – proposée par Doganaksoy, Faltin et Tucker [2] - qui ont permis à Mason, Tracy et Young [80-82] d'élaborer une méthode plus complexe et performante. Comme nous avons vu dans les paragraphes précédents, les techniques statistiques de régression linéaire multiple, d'analyse en composantes principales, sont les plus souvent utilisées pour localiser la source d'un dérèglement sur les cartes multivariées. Ces deux techniques sont très complexes et peuvent donner lieu à beaucoup d'interprétations. Les statisticiens ont toujours trouvé des cas pratiques particuliers qui sont solutionnés d'une façon plus adéquate avec une méthode ou avec

une autre. Parmi les techniques basées sur la régression multiple, la méthode MYT de décomposition de la statistique T^2 est, sans aucun doute la plus performante, si l'hypothèse de distribution normale multiple pour les variables est respectée. Dans le cas des processus en cascade la méthode de sélection de la cause de Zhang donne également de bons résultats. Il nous semble utile de préciser que les techniques de régression multiple ne sont pas très efficaces quand la corrélation entre les variables est forte. Dans ce cas, la méthode d'analyse en composantes principales (ACP) et la méthode des projections dans les structures latentes (PLS) donnent de meilleurs résultats. En plus, l'analyse en composantes principales a l'avantage de nécessiter un nombre réduit de cartes de contrôle ce qui facilite l'analyse.

Malgré la diversité des concepts statistiques utilisés, presque toutes les méthodes présentées dans ce chapitre ont un point en commun. Elles sont des méthodes "off-line", c'est-à-dire qu'elles ne peuvent pas être utilisées en temps réel pour le diagnostic du processus. Quand une situation hors contrôle est détectée dans un processus multivarié, une analyse complexe et détaillée doit être effectuée afin de déterminer la variable ou les variables responsable(s) de cette situation. Généralement, cette analyse ne peut pas être effectuée en temps réel.

La plupart des méthodes font appel à des représentations graphiques des caractéristiques univariées indépendantes comme les résidus ou les composantes principales par exemple. Ces variables indépendantes sont ensuite représentées sur des cartes de contrôle de type Shewhart. Quand le nombre de variables est très élevé il devient difficile d'analyser simultanément toutes ces cartes.

Dans la plupart des cas, l'opérateur doit être formé avant d'utiliser la méthode et ses conclusions dépendent, dans beaucoup de cas, de ses connaissances statistiques et parfois de son intuition. Après avoir détecté la source de la situation hors contrôle, l'expérience acquise par l'analyse du processus n'est pas valorisée d'une manière efficace, c'est-à-dire qu'une nouvelle analyse sera effectuée pour chaque situation hors contrôle comme si aucune connaissance du processus n'était disponible.

Nous proposons une nouvelle méthodologie de contrôle de processus basée sur le filtrage numérique et sur l'analyse discriminante afin de remédier à ces inconvénients. Nous allons décrire plus en détail cette méthodologie dans le chapitre 3.

Chapitre 3

- 3.1 Introduction
- 3.2 Analyse Discriminante
- 3.3 Filtrage Numérique
- 3.4 Méthodologie FNAD
- 3.5 Application industrielle
- 3.6 Conclusion

3.1 Introduction

Nous avons présenté dans le chapitre 2 un panorama non-exhaustif des méthodes de détection de la variable responsable d'une situation hors-contrôle dans le cas multivarié. Nous avons analysé les avantages et les inconvénients de ces méthodes dans le paragraphe 2.4.

Dans ce chapitre nous proposons une nouvelle méthodologie pour la détection de la source d'un dérèglement dans un processus multivarié. Notre méthode repose principalement sur l'utilisation de l'analyse discriminante (AD) et du filtrage numérique de signal (FN). Par conséquent, nous aborderons au début de ce chapitre quelques notions de base concernant l'analyse discriminante. Les deux principaux aspects théoriques de la méthode, l'aspect descriptif et l'aspect décisionnel, seront exposés. Quelques remarques concernant les caractéristiques qualitatives des règles de classement seront énoncées. Dans le deuxième paragraphe de ce chapitre, nous allons donner une définition du signal numérique et du filtre numérique de signal. Ensuite, nous allons passer en revue les différents types de filtres numériques et nous allons préciser les équations récurrentes qui les caractérisent. Finalement, dans le dernier paragraphe, nous allons donner une description détaillée de la méthodologie de contrôle FNAD (Filtrage Numérique et Analyse Discriminante) et un exemple d'application pratique basé sur des données simulées sera analysé.

3.2 Analyse discriminante

On désigne sous le nom *d'analyse discriminante* une famille de techniques destinées à classer (affecter à des classes préexistantes) des individus caractérisés par un certain nombre de variables numériques ou nominales (Lebart et al. [70]). Les données consistent en n observations (ou individus) réparties en q classes et décrites par p variables explicatives (ou prédicteurs). On distingue classiquement deux aspects en analyse discriminante:

- ✓ **descriptif**: chercher les combinaisons linéaires des variables explicatives permettant de séparer le mieux possible les q classes existantes et de donner ainsi une représentation graphique qui rende compte au mieux de cette séparation.
- ✓ **décisionnel**: décider dans quelle classe il faut affecter un nouvel individu pour lequel on connaît les valeurs des prédicteurs. C'est le type de problème auquel on se réfère le plus souvent lorsqu'on parle de discrimination: il s'agit d'identifier la classe de l'individu. Il ne faut pas confondre ce problème de classement avec un problème de classification.

3.2.1 Formulation du problème et notations

On dispose de n individus, ou observations, décrits par un ensemble de p variables (v_1, v_2, \dots, v_p) et répartis en q classes définies a priori par la variable nominale y (figure III.1). L'analyse discriminante se propose, dans un premier temps, de séparer aux mieux les q classes à l'aide de ces p variables explicatives. Dans un deuxième temps, elle cherche à résoudre le problème de l'affectation d'individus nouveaux, caractérisés également par les p variables, à certaines classes déjà identifiées sur l'échantillon de n individus (appelé échantillon d'apprentissage).

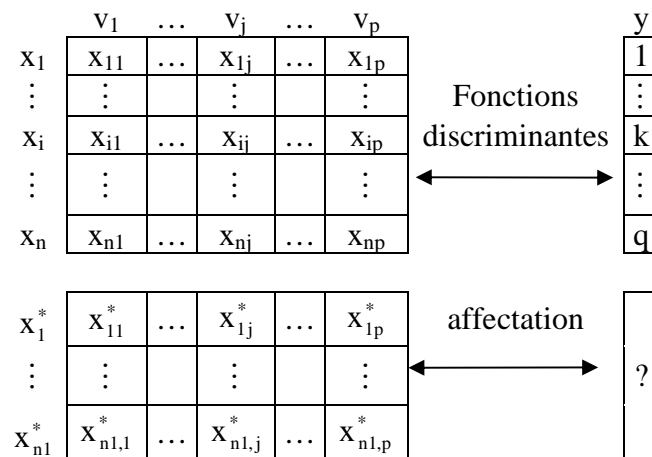


figure III.1 - Principe de l'analyse discriminante

Les n individus x_i de l'échantillon d'apprentissage constituent un nuage de points \mathbf{Y} dans \mathbf{R}^p divisé en q sous-nuages $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_q$ ayant leurs centres de gravité $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_k$ (figure III.2).

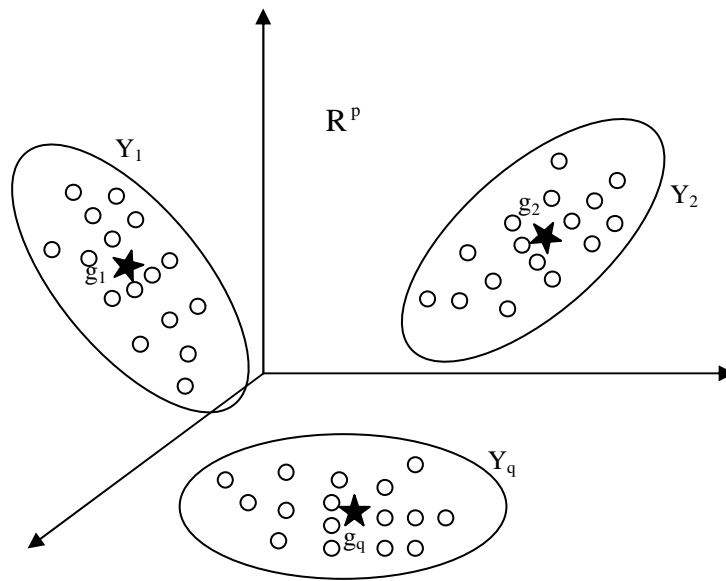


figure III.2 - Nuage des points dans R^p

Si les n individus x_i sont affectés des poids $m_{x_1}, m_{x_2}, \dots, m_{x_n}$ alors les poids $m_{Y_1}, m_{Y_2}, \dots, m_{Y_q}$ de chaque sous-nuage sont:

$$m_{Y_j} = \sum_{x_i \in Y_j} m_{x_i} \quad (3.1)$$

La masse totale du nuage Y vaut ici: $m_Y = \sum_{i=1}^n m_{x_i} = 1$.

Les centres de gravités (moyennes) de chaque sous-nuage sont calculés à l'aide de la formule:

$$g_j = \frac{1}{m_{Y_j}} \sum_{x_i \in Y_j} m_{x_i} x_i \quad (3.2)$$

Par conséquent, le centre de gravité du nuage Y est: $g = \sum_{j=1}^q m_{Y_j} g_j$

L'inertie (variance) de chaque sous-nuage est:

$$V_j = \frac{1}{m_{Y_j}} \sum_{x_i \in Y_j} m_{x_i} (x_i - g_j)(x_i - g_j)^T \quad (3.3)$$

La *matrice de variance intra-classe* (V_{intra}) est définie comme la moyenne des matrices V_j :

$$V_{\text{intra}} = \sum_{j=1}^q m_{Y_j} V_j \quad (3.4)$$

et la *matrice de variance inter-classe* (V_{inter}) comme la matrice de variance de q centres

de gravité affectés des poids m_{Y_j} :

$$V_{\text{inter}} = \sum_{j=1}^q m_{Y_j} (\mathbf{g}_j - \mathbf{g})(\mathbf{g}_j - \mathbf{g})^T \quad (3.5)$$

L'inertie totale du nuage Y par rapport à son centre de gravité \mathbf{g} est:

$$V_{\text{totale}} = \sum_{i=1}^n m_{x_i} (\mathbf{x}_i - \mathbf{g})(\mathbf{x}_i - \mathbf{g})^T \quad (3.6)$$

Un théorème classique en mécanique, le théorème de Huygens, appliqué au nuage Y, nous permet d'écrire:

$$V_{\text{totale}} = V_{\text{intra}} + V_{\text{inter}} \quad (3.7)$$

exprimant le fait que l'inertie d'un ensemble de sous-nuages de points, par rapport à son centre de gravité, est égale à la somme des inerties de chacun des sous-nuages par rapport à leurs centres de gravité respectifs augmentée de l'inertie, par rapport au centre de gravité global, de l'ensemble des masses de chaque sous-nuage supposées concentrées aux centres de gravité respectifs.

Nous supposons désormais que $\mathbf{g} = \mathbf{0}$, c'est-à-dire que les variables explicatives sont centrées. Dans le cas où $m_{x_i} = \frac{1}{n}$, les équations (3.2), (3.4) et (3.5) se simplifient et en introduisant les effectifs n_1, n_2, \dots, n_q des q sous-nuages, on a:

$$\mathbf{g}_j = \frac{1}{n_j} \sum_{x_i \in Y_j} \mathbf{x}_i, \quad V_{\text{intra}} = \frac{1}{n} \sum_{j=1}^q n_j V_j, \quad V_{\text{inter}} = \frac{1}{n} \sum_{j=1}^q n_j \mathbf{g}_j \mathbf{g}_j^T \quad (3.8)$$

Si on considère l'ensemble de p variables, on définit une matrice de covariance dont les termes diagonaux seront les covariances relatives à un couple de variables. De même que nous avons défini trois notions de variance, nous allons définir trois types de matrices de covariance:

- ✓ matrice de covariance totale: $T = \frac{1}{n} \sum_{k=1}^q \sum_{x_i \in Y_k} \mathbf{x}_i^T \mathbf{x}_i$
- ✓ matrice de covariance intra-classe: $W = \frac{1}{n} \sum_{k=1}^q \sum_{x_i \in Y_k} n_k (\mathbf{x}_i - \mathbf{g}_k)^T (\mathbf{x}_i - \mathbf{g}_k)$
- ✓ matrice de covariance inter-classe: $B = \sum_{k=1}^q \frac{n_k}{n} \mathbf{g}_k^T \mathbf{g}_k$

Le théorème de Huygens nous permet encore d'écrire l'égalité matricielle:

$$T = W + B \quad (3.9)$$

Si on projette l'équation (3.9) sur une forme linéaire (\mathbf{u}) définie sur \mathbb{R}^p , alors on obtient:

$$\mathbf{uTu}^T = \mathbf{uWu}^T + \mathbf{uBu}^T \quad (3.10)$$

Dans le prochain paragraphe on cherche les formes linéaires \mathbf{u} (nommées aussi axes factoriels discriminants) qui permettent de discriminer au mieux l'ensemble de classes d'individus.

3.2.2 Axes factoriels discriminants. Métrique discriminante

Nous voudrions ici trouver l'axe factoriel, ou la forme linéaire, \mathbf{u} qui discrimine au mieux l'ensemble des classes d'individus. La discrimination sera d'autant plus aisée que les classes sont éloignées les unes des autres (variance inter-classe grande) et que les individus d'une même classe seront proches (variance intra-classe petite).

Dès lors, le premier axe factoriel discriminant \mathbf{u}_1 sera l'élément \mathbf{u} qui maximise le rapport de la variance inter-classe de \mathbf{u} à la variance intra-classe de \mathbf{u} :

$$\underset{(u)}{\text{maximiser}} \left(\frac{\mathbf{uBu}^T}{\mathbf{uWu}^T} \right) \quad (3.11)$$

De fait, puisque: $\mathbf{uTu}^T = \mathbf{uWu}^T + \mathbf{uBu}^T$, voir l'équation (3.10) on a les équivalences:

$$\begin{aligned} \underset{(u)}{\text{maximiser}} \left(\frac{\mathbf{uBu}^T}{\mathbf{uWu}^T} \right) &\leftrightarrow \underset{(u)}{\text{minimiser}} \left(\frac{\mathbf{uWu}^T}{\mathbf{uBu}^T} \right) \leftrightarrow \underset{(u)}{\text{minimiser}} \left[\left(\frac{\mathbf{uWu}^T}{\mathbf{uBu}^T} \right) + 1 \right] \leftrightarrow \\ &\leftrightarrow \underset{(u)}{\text{minimiser}} \left(\frac{\mathbf{uTu}^T}{\mathbf{uBu}^T} \right) \leftrightarrow \underset{(u)}{\text{maximiser}} \left(\frac{\mathbf{uBu}^T}{\mathbf{uTu}^T} \right) \end{aligned}$$

Nous sommes donc ramenés à résoudre le problème équivalent:

$$\underset{(u)}{\text{maximiser}} \left(\frac{\mathbf{uBu}^T}{\mathbf{uTu}^T} \right) \quad (3.12)$$

Il est possible de montrer, voir l'annexe A.4, que le premier axe factoriel discriminant \mathbf{u}_1 est le vecteur propre de $T^{-1}B$ correspondant à la plus grande valeur propre λ_1 . De même, on définit le deuxième axe factoriel discriminant \mathbf{u}_2 comme le deuxième vecteur propre de $T^{-1}B$ orthogonal à \mathbf{u}_1 . Il constitue le meilleur facteur discriminant indépendant du premier.

Il se pose alors la question du nombre d'axes factoriels significativement discriminants que l'on peut extraire.

Tout d'abord, remarquons que dans la plupart des problèmes de discrimination, le nombre total d'individus \mathbf{n} est supérieur au nombre de variables \mathbf{p} , lui même supérieur au nombre de classes \mathbf{q} . Par conséquent, les individus x_i engendrent l'espace R^p , tandis que les centres de gravité de chacune des \mathbf{q} classes engendrent un sous espace \mathbf{V} de dimension $\mathbf{q}-1$. Les formes linéaires discriminantes peuvent se définir géométriquement de la manière suivante: dans \mathbf{V} , muni de la métrique induite par T^{-1} , on détermine les

vecteurs directeurs $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{q-1}$ des axes principaux d'inertie du système des points $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_q$ affectés des masses $m_{Y_1}, m_{Y_2}, \dots, m_{Y_q}$; la forme linéaire \mathbf{u}_1 sera $u_1 = \mathbf{f}_1^T \mathbf{T}^{-1}$.

Sous l'hypothèse $n > p > q$, il y a donc exactement $q-1$ vecteurs propres de $\mathbf{T}^{-1}\mathbf{B}$, c'est-à-dire $q-1$ axes factoriels discriminants. Rao [101] propose un test pour vérifier la signification statistique des axes factoriels discriminants.

On appelle "pouvoir discriminant" de l'axe factoriel \mathbf{u}_1 la quantité λ_1 . Selon l'équation (3.10) toutes les valeurs propres λ de $\mathbf{T}^{-1}\mathbf{B}$ sont positives et inférieures à 1. Si l'on considère la variance totale selon la direction définie par \mathbf{u}_1 , ce pouvoir discriminant est la fraction de cette variance due à la différence entre les classes.

Remarque:

Nous avons montré que les axes factoriels discriminants sont les vecteurs propres de $\mathbf{T}^{-1}\mathbf{B}$. La solution de l'équation (3.11) donne les vecteurs propres de $\mathbf{W}^{-1}\mathbf{B}$. Il s'ensuit que les vecteurs propres de $\mathbf{T}^{-1}\mathbf{B}$ et de $\mathbf{W}^{-1}\mathbf{B}$ sont nécessairement les mêmes. Seules, par contre, diffèrent les valeurs propres correspondantes. Nous allons montrer que si λ et \mathbf{u}^T sont la valeur et le vecteur propre de $\mathbf{T}^{-1}\mathbf{B}$, alors $\frac{\lambda}{1-\lambda}$ et \mathbf{u}^T sont respectivement la valeur et le vecteur propre de $\mathbf{W}^{-1}\mathbf{B}$ (pourvu que \mathbf{W} soit inversible):

$\mathbf{T}^{-1}\mathbf{B}\mathbf{u}^T = \lambda\mathbf{u}^T \Rightarrow \mathbf{B}\mathbf{u}^T = \lambda\mathbf{T}\mathbf{u}^T$ et comme selon l'équation (3.9) $\mathbf{T} = \mathbf{B} + \mathbf{W}$ on obtient:

$$\mathbf{B}\mathbf{u}^T = \lambda\mathbf{B}\mathbf{u}^T + \lambda\mathbf{W}\mathbf{u}^T \Rightarrow \mathbf{B}\mathbf{u}^T = \frac{\lambda}{1-\lambda}\mathbf{W}\mathbf{u}^T \Rightarrow \mathbf{W}^{-1}\mathbf{B}\mathbf{u}^T = \frac{\lambda}{1-\lambda}\mathbf{u}^T$$

L'utilisation de \mathbf{T}^{-1} ou de \mathbf{W}^{-1} comme métrique est donc indifférente. La métrique \mathbf{W}^{-1} est appelée "métrique de Mahalanobis".

3.2.3 Règles d'affectation ou de classement

Dans le précédent paragraphe nous avons présenté la discrimination à but descriptif. Nous allons dans ce paragraphe passer en revue l'aspect décisionnel de l'analyse discriminante qui consiste à déterminer la classe d'appartenance d'un nouvel individu x_i^* pour lequel on connaît les valeurs des prédicteurs $x_{i1}^*, x_{i2}^*, \dots, x_{ip}^*$. On peut mentionner, ici, l'existence de deux grandes catégories de règles d'affectation:

- ✓ les règles géométriques d'affectation;
- ✓ les règles probabilistes d'affectation;

3.2.3.1 Règles géométriques d'affectation

Une des règles géométriques d'affectation les plus connues est la règle de Mahalanobis-Fisher. Cette règle consiste à calculer la distance entre le nouvel individu et les q centres des classes et à affecter l'individu à la classe ayant le centre de gravité le plus proche. Les distances sont calculées à l'aide de la métrique \mathbf{T}^{-1} (où \mathbf{T} est la matrice

de covariance totale). L'avantage principal de cette règle d'affectation est la linéarité. Elle consiste à calculer pour chaque individu x_i^* , q distances qui représentent des fonctions linéaires de ses coordonnées, et à en chercher la valeur minimale. Un des inconvénients majeurs de la règle Mahalanobis-Fisher est qu'elle conduit à des affectations incorrectes lorsque les dispersions des groupes sont très différentes. Rien ne justifie alors l'usage de la même métrique pour les différents groupes.

Diverses solutions utilisant des métriques locales ont été proposées. Une des plus connues est l'approche de Sebestyen. Dans cette approche deux indicateurs sont définis:

- ✓ *la similitude* entre un nouvel individu et une classe (définie comme étant la distance moyenne au carré entre l'individu et les membres de la classe);
- ✓ *l'agrégation* d'une classe (la distance moyenne entre les individus d'une classe);

On cherche ensuite à trouver une métrique M qui permettra la minimisation de la distance entre les individus appartenant à la même classe, tout en respectant en même temps une certaine contrainte de normalisation. Il est possible de démontrer que cette métrique est de la forme suivante:

$$M = |\Sigma_k|^{1/P} \Sigma_k^{-1} \quad (3.13)$$

où $|\Sigma_k|$ est le déterminant de la matrice variance-covariance de la classe k .

La métrique M définie dans l'équation (3.13) est équivalente à une transformation linéaire sur les variables initiales. Cette transformation linéaire est le produit d'une rotation (définie par la matrice des vecteurs propres de Σ_k) et d'une transformation diagonale relative aux nouvelles variables obtenues après rotation. La méthode est étroitement liée à la méthode d'analyse en composantes principales (ACP), mais dans l'analyse discriminante (AD) se sont les dernières composantes principales (de faible variance) qui sont utilisées.

3.2.3.2 Règles probabilistes d'affectation

Une des règles probabilistes d'affectation les plus utilisées est la *règle bayésienne*. Si on suppose que les q groupes, sont en proportion de p_1, p_2, \dots, p_q dans la population totale et que la distribution de probabilité du vecteur d'observations est donnée pour chaque groupe k par une fonction de densité de probabilité $f_k(x)$, alors la probabilité qu'un point x appartienne au groupe k est donnée par la formule de Bayes:

$$P(Y_k|x) = \frac{p_k f_k(x)}{\sum_{k=1}^q p_k f_k(x)} \quad (3.14)$$

La règle bayésienne consiste à affecter l'observation x au groupe qui a la probabilité *a posteriori* maximale. Les dénominateurs étant les mêmes pour les q groupes on doit donc chercher le maximum de: $p_k f_k(x)$. Il est donc nécessaire de connaître ou d'estimer $f_k(x)$. La fonction densité de probabilité peut s'estimer à l'aide des

méthodes non-paramétriques (comme la méthode de noyau ou la méthode des k plus proches voisins) qui ne font pas d'hypothèse spécifique sur la famille de loi de probabilité ou à l'aide des *méthodes paramétriques* qui à partir d'une famille paramétrée de lois de probabilité utilisent l'échantillon pour estimer les paramètres (le cas de la distribution p-normale est le plus classique).

3.2.4 Qualité des règles de classement

La mesure de la qualité d'une discrimination se fait à partir du pourcentage d'observations bien classées pour chaque classe, et du pourcentage global des observations bien classées. Si on calcule le pourcentage d'observations bien classées sur l'échantillon d'apprentissage, la qualité de la discrimination est surestimée. Le pourcentage d'observations bien classées augmente avec le nombre des paramètres du modèle, et il peut être excellent si le nombre des paramètres est considérable, sans pour cela assurer que le modèle permet de réaliser une prévision correcte. Le pourcentage de mal classées dans ces conditions est appelé taux d'erreur apparent (ou de resubstitution). Plusieurs méthodes existent pour le calcul du pourcentage d'observations bien classées. La *méthode d'échantillons test* consiste à réaliser la discrimination sur une partie de l'échantillon d'apprentissage (environ 80%) et le test sur les autres. On peut améliorer le calcul du taux d'erreur en divisant l'échantillon d'apprentissage en m groupes et en calculant les axes discriminants sur $m-1$ groupes. Le taux d'erreur est calculé sur le groupe restant. L'avantage est que l'on peut calculer ainsi un nombre de m taux d'erreur et finalement un taux d'erreur moyen sur un échantillon de la taille de l'échantillon d'apprentissage. Plus m est proche de n , plus on se rapproche de la situation réelle de classement. La *validation croisée* correspond au cas $m = n$, autrement dit, au cas pour lequel on effectue n discriminations excluant à chaque fois une observation (l'estimateur "leave-one-out"). Cet estimateur, bien qu'il ne soit pas biaisé, a une grande variance pour des petits échantillons. Cette variance peut être suffisamment élevée pour que les approches ayant un biais légèrement plus grand, mais avec une variance plus faible, surpasse l'estimée "leave-one-out" de l'erreur de classification.

La méthode "bootstrap" échantillonne les données avec remplacement pour donner de nouveaux ensembles de données. Cette technique est très coûteuse en temps de calcul. Malheureusement, les études de Chernick et al. [12], Jain et al. [57] ont mis en évidence qu'il n'existe pas un seul estimateur bootstrap utilisable pour tous les classifieurs et dans tous les domaines d'utilisation.

Des estimateurs composés qui combinent les différents estimateurs suivant le domaine où il sont les plus performants ont été proposés. L'estimée de la probabilité d'erreur de classification peut se calculer comme étant la moyenne de l'estimée "leave-one-out" et de l'estimée par resubstitution [102]

Devijver et Kittler [21] proposent un intervalle d'estimation paramétrique de l'erreur de classification basé sur la loi normale. Pour sa part, Fukunaga [34] utilise l'estimation de l'erreur de classification par la méthode "leave-one-out" et par la méthode de resubstitution pour obtenir un intervalle non paramétrique de la probabilité d'erreur de classification.

3.3 Filtrage numérique

Dans ce paragraphe, nous allons définir le signal numérique et le filtre numérique de signal. Ensuite, nous allons passer en revue les différents types de filtres numériques et nous allons préciser les équations récurrentes les caractérisant.

3.3.1 Introduction

Un signal est qualifié de *numérique* s'il transporte une information qui est une simple suite de chiffres. Il est donc discontinu dans le temps (les chiffres se suivent les uns après les autres) et pour l'une de ses caractéristiques (amplitude, fréquence ou phase) ne pouvant prendre qu'un nombre fini d'états. C'est l'opération de mesure qui fait correspondre à la valeur instantanée de la grandeur physique un nombre dont la précision dépend du dispositif de mesure employé.

Pour transformer un signal analogique continu $x(t)$ en signal numérique, deux opérations sont nécessaires :

- ✓ prélever périodiquement des valeurs successives du signal, c'est l'opération *d'échantillonnage*; si T_e est l'intervalle d'échantillonnage, le résultat de l'échantillonnage est une suite de valeurs $x(kT_e)$;
- ✓ remplacer $x(kT_e)$ par un nombre qui est le résultat de sa mesure avec une certaine unité de mesure; ce nombre est ensuite utilisé pour modifier, via un code, la grandeur physique constituant le signal; cette seconde opération est la *conversion analogique-numérique* ou *quantification*.

On appelle *filtrage numérique de fréquences* ou *filtrage numérique linéaire* toute combinaison linéaire des échantillons de l'entrée et de la sortie d'un système échantillonné (système numérique dans la quasi-totalité des cas).

Si le signal de l'entrée a été échantillonné à la fréquence f_e (soit un intervalle d'échantillonnage T_e), l'équation linéaire de filtrage la plus générale est:

$$b_0 y(kT_e) = \sum_{i=0}^n a_i x(kT_e - iT_e) + \sum_{j=1}^m b_j y(kT_e - jT_e) \quad (3.15)$$

où b_0 , a_i et b_j sont des constantes appelées aussi *coefficients du filtre*. Afin de simplifier l'écriture de l'équation précédente, on utilise le plus souvent la notation:

$$b_0 y_k = \sum_{i=0}^n a_i x_{k-i} + \sum_{j=1}^m b_j y_{k-j} \quad (3.16)$$

où les indices $k-i$ ou $k-j$ signifient des échantillons du signal aux instants $kT_e - iT_e$ et respectivement $kT_e - jT_e$. D'une façon générale, pour simplifier encore plus, on prend $b_0 = 1$.

En utilisant la *transformée en z* et notamment la *propriété de translation* de la transformée en z ($TZ(x_{t-\lambda}) = X(z)z^{-\lambda}$ ou autrement, si l'on décale de λT_e tous les échantillons du signal numérique, sa transformée en z est multipliée par $z^{-\lambda}$) l'équation (3.16) peut s'écrire sous la forme suivante:

$$Y(z) = \sum_{i=0}^n a_i X(z) z^{-i} + \sum_{j=1}^m b_j Y(z) z^{-j} = X(z) \sum_{i=0}^n a_i z^{-i} + Y(z) \sum_{j=1}^m b_j z^{-j}$$

d'où on peut facilement en déduire la fonction de transfert en z du filtre numérique:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{i=0}^n a_i z^{-i}}{1 - \sum_{j=1}^m b_j z^{-j}} \quad (3.17)$$

L'équation (3.17) montre que la fonction de transfert d'un filtre numérique linéaire a obligatoirement la forme d'un rapport de deux polynômes algébriques en z .

La conception des filtres numériques consiste à trouver les coefficients du filtre afin d'assurer une certaine fonction souhaitée. Pour ce faire, il existe plusieurs méthodes qui, en partant de l'expression de la fonction de transfert analogique d'un filtre, permettent d'estimer les coefficients du filtre numérique équivalent. Les deux méthodes les plus utilisées sont:

- ✓ la méthode de transposition des pôles associée à la méthode de l'équivalence de la dérivation;
- ✓ la méthode de la transformation bilinéaire.

3.3.2 Types de filtres numériques

La fonction de transfert d'un filtre numérique présente les caractéristiques suivantes (figure III.3):

- **bande passante** $[0, \omega_p]$ – le domaine des fréquences dans lequel le pouvoir spectral du signal d'entrée passe à travers le filtre avec un gain approximatif égal à 1;
- **bande d'arrêt** $[\omega_s, \pi]$ – le domaine des fréquences dans lequel le pouvoir spectral du signal est atténué jusqu'à l'élimination;
- **bande de transition** $[\omega_p, \omega_s]$ – le domaine de fréquences compris entre la bande de passage et la bande d'arrêt;

Selon la forme de la fonction de transfert, les filtres numériques sont considérés comme faisant partie d'une de ces quatre catégories de base: *filtre passe-bas* (qui ne laisse passer que les basses fréquences jusqu'à une limite fixée), *filtre passe-haut* (qui ne laisse passer que les hautes fréquences au-delà d'une valeur déterminée), *filtre passe-bande* (qui laisse passer une gamme de fréquences située entre des valeurs déterminées) et *filtre coupe-bande* (qui laisse passer toutes les fréquences sauf celles d'une bande déterminée).

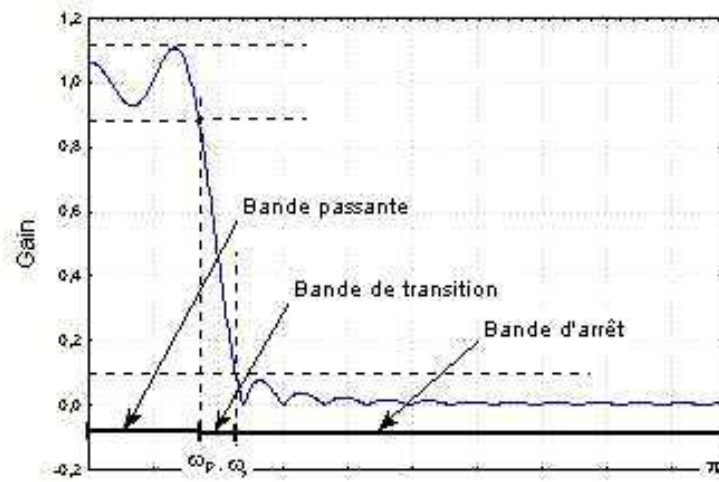


figure III.3 - Caractéristiques de la fonction de transfert d'un filtre numérique

L'équation (3.17) représente la fonction de transfert en z la plus générale pour un filtre linéaire. Un tel filtre est appelé récursif parce que, en vertu de l'équation (3.16), le calcul de la valeur y_k de la sortie nécessite non seulement les valeurs antérieures de l'entrée (x_{k-i}), mais aussi les valeurs antérieures de la sortie (y_{k-j}).

Si dans l'équation (3.17) tous les coefficients a_i sont nuls sauf a_0 la fonction de transfert en z du filtre devient:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{a_0}{1 - \sum_{j=1}^m b_j z^{-j}}$$

d'où:

$$Y(z) = a_0 X(z) + \sum_{j=1}^m b_j Y(z) z^{-j} = a_0 X(z) + b_1 Y(z) z^{-1} + b_2 Y(z) z^{-2} + \dots + b_m Y(z) z^{-m}$$

ou, en passant dans le domaine temporel:

$$y_t = a_0 x_t + b_1 y_{t-1} + b_2 y_{t-2} + \dots + b_m y_{t-m} \quad (3.18)$$

L'équation (3.18) montre que la sortie du filtre y ne dépend que de la valeur actuelle de l'entrée x et des m valeurs précédentes de la sortie. Un tel filtre est appelé **autorégressif (AR)**.

Si dans l'équation (3.17) tous les coefficients b sont nuls, la fonction de transfert en z du filtre devient:

$$H(z) = \frac{Y(z)}{X(z)} = \sum_{i=0}^n a_i z^{-i}$$

La fonction de transfert ne comporte plus de dénominateur et on a :

$$Y(z) = \sum_{i=0}^n a_i X(z) z^{-i}$$

ou, en revenant dans le domaine temporel:

$$y_t = \sum_{i=0}^n a_i x_{t-i} \quad (3.19)$$

On peut remarquer dans l'équation (3.19) que la sortie y du filtre ne dépend que des entrées précédentes. Un tel filtre est appelé **moyenne mobile (MM)** ou *moving average (MA)* en anglais.

La relation (3.19) ne représente rien d'autre que l'équation de convolution numérique dans le domaine temporel. Les coefficients a_i sont les valeurs successives de la réponse impulsionnelle du filtre numérique. Du fait que le nombre de termes de la somme de la relation est limité, les filtres MM sont souvent appelés filtres à réponse impulsionnelle finie (de durée finie), **filtres RIF** (ou *FIR* - "*Finite Impulse Response*" en anglais).

Par contre, pour une fonction de transfert en z décrite par l'équation (3.17), si l'on opère la division du numérateur par le polynôme algébrique en z du dénominateur, on trouvera, en général, une infinité de termes. Des tels filtres sont appelés filtres à réponse impulsionnelle infinie (de durée infinie), **filtres RII** (ou *IIR* - "*Infinite Impulse Response*" en anglais).

Le filtre le plus général, voir l'équation (3.17), peut être considéré comme l'association en cascade d'une fonction de transfert autorégressive et d'une fonction de transfert à moyenne mobile, d'où la dénomination **ARMA** (*AutoRegressive and Moving Average*) qui les désigne.

3.4 Méthode FNAD (Filtrage Numérique et Analyse Discriminante)

3.4.1 Introduction

Nous avons analysé dans le chapitre 2 les différentes méthodes de détection de la source d'un dérèglement dans le cas multivarié et nous avons mis en évidence leurs avantages et leurs inconvénients. Peu de ces méthodes tiennent compte de la connaissance existante au sujet du processus analysé et leur application s'effectue généralement comme si aucune information préliminaire n'était disponible. Dans le meilleur des cas, les connaissances techniques concernant le processus de fabrication sont utilisées seulement comme un appui pour confirmer les résultats statistiques obtenus et pour aider à l'interprétation de ces résultats. Pire encore, après avoir dépensé beaucoup d'effort, d'énergie et de temps, en dépouillant et en analysant les données, l'expérience ainsi acquise n'est pas valorisée d'une manière efficace. Cela veut dire qu'à chaque fois qu'une situation hors-contrôle est détectée les analyses devront être reprises de le début.

Dans ce paragraphe, nous proposons une nouvelle méthodologie de contrôle en temps réel permettant la détection de la source d'un dérèglement dans le cas multivarié. Notre méthodologie, nommée **FNAD**, est basée principalement sur l'utilisation du **F**iltrage **N**umérique et de l'**A**nalyse **D**iscriminante. L'objectif de la méthode est double: diagnostiquer l'état de fonctionnement d'un processus multivarié et préciser dans le cas d'un fonctionnement anormal les actions nécessaires à effectuer afin de remédier à la situation. Il faut bien préciser ici que nous voulons concevoir une méthodologie de contrôle et de suivi de processus intégrant les connaissances des industriels à l'égard de leur processus de fabrication. Cette expérience acquise dans le temps et cette source de savoir faire nous les considérons comme primordiales dans toute démarche d'amélioration de la qualité et de résolution de problème. Nous voulons, par conséquent, gagner en efficacité par l'intégration de ces informations au cœur de notre système de détection.

Le schéma bloc fonctionnel de la méthodologie FNAD, présenté dans la figure III.4, comporte trois modules principaux:

- ✓ le module d'analyse "on-line" (ou chaîne directe) qui est composé à son tour de trois sous-modules: le module de filtrage numérique, le module de rejet et le module d'analyse discriminante;
- ✓ le module "bibliothèque des classes" qui résume dans un certain nombre de paramètres l'information concernant les caractéristiques liées au fonctionnement normal et au fonctionnement anormal du processus;
- ✓ le module d'analyse "off-line" qui est structuré autour d'un sous-module de classification; ce module est utilisé comme un outil d'analyse exploratoire des données dont le but principal est l'identification des différentes classes de fonctionnement du processus;

Nous allons expliquer le fonctionnement global du schéma de la figure III.4 et ensuite nous allons traiter plus en détail chaque bloc séparément.

Le module d'analyse "on-line" est utilisé pour la détection en temps réel des dérèglements et pour le diagnostic. A l'entrée du module les données multivariées passent par un module de filtrage numérique qui a pour but de réduire leur variabilité. Le diagnostic est réalisé à l'aide du module d'analyse discriminante. A la sortie de ce module chaque observation est classée dans une des classes existantes (déjà identifiées par une analyse "off-line" d'un ensemble de données historiques du processus ou suite à une expérimentation visant la mise en évidence de ces classes). Le module de rejet est optionnel (il est utile surtout quand toutes les classes ou modes de fonctionnement n'ont pas été identifiés lors de l'analyse "off-line"). Son rôle est de ne pas permettre aux observations appartenant à un nouveau mode de fonctionnement d'être classées par erreur parmi les observations d'une des classes existantes. Ces observations seront déposées temporairement dans une mémoire tampon et participeront à la mise à jour de la bibliothèque de classes. Dans cette bibliothèque, chaque classe regroupe les observations caractérisant un mode de fonctionnement. Chaque situation hors-contrôle ayant à l'origine le dérèglement d'une ou plusieurs variables de processus sera représentée par une classe de fonctionnement anormal. Les caractéristiques de chaque classe seront bien déterminées et précisées à cette étape de l'analyse de l'évolution historique du processus. Ensuite, au fur et à mesure du déroulement du processus, chaque classe s'enrichira avec des nouvelles observations et elle sera ré-actualisée en permanence. Si le nombre d'observations constituant les classes devient très important il se peut que

l'analyse en temps réel soit ralentie à cause du volume de calculs nécessaires. Dans ce cas, nous proposons une analyse approfondie de chaque classe afin de déterminer leur distribution probabiliste et d'identifier les paramètres associés. Ainsi, au lieu de conserver chaque observation, les classes pourront être décrites par leurs moyennes et leurs matrices de variance-covariance, par exemple, ce qui est nettement plus avantageux pour le calcul et pour l'espace de stockage nécessaire. Cette bibliothèque de classes est le résultat d'une analyse approfondie de l'existant, analyse qui est effectuée dans le module d'analyse "off-line". Ce module est construit autour d'un module de classification mixte (classification hiérarchique et non-hiérarchique) et a pour but l'identification des différents régimes de fonctionnement du processus sur la base des données enregistrées dans une base de données utilisateur. On utilise, également, un module de filtrage numérique identique à celui qui se trouve dans le module d'analyse "on-line". Son rôle est de réduire la variabilité des données afin de minimiser les erreurs de classification. Dans le module de pré-traitement, diverses opérations sont effectuées sur les données, dans le but de rendre la base de données exploitable. Le module de sélection des variables permet la réduction de la dimension de l'espace de représentation des données. Il aide à déterminer l'ensemble des variables mettant en évidence les différents régimes de fonctionnement enregistrés dans la base de données utilisateur.

Après cette présentation, très succincte, du schéma de fonctionnement de la méthode FNAD, nous allons dans les paragraphes suivants focaliser notre attention sur chacun des ces modules. Nous allons décrire plus en détail leur rôle dans la logique de contrôle que nous proposons, les résultats obtenus et, finalement, nous allons souligner leurs avantages et leurs inconvénients.

Mais, avant d'entrer dans les détails du schéma de la figure III.4, il est nécessaire de faire quelques remarques et précisions au sujet des dérèglages que l'on veut détecter par la méthode FNAD. Nous allons analyser dans les deux paragraphes suivants leur nature et leurs caractéristiques. Une analyse détaillée sera effectuée dans le domaine temporel et spectral. Cette analyse nous servira, par la suite entre autres, à expliquer le choix de la méthode de filtrage utilisée dans le cadre de la méthode FNAD.

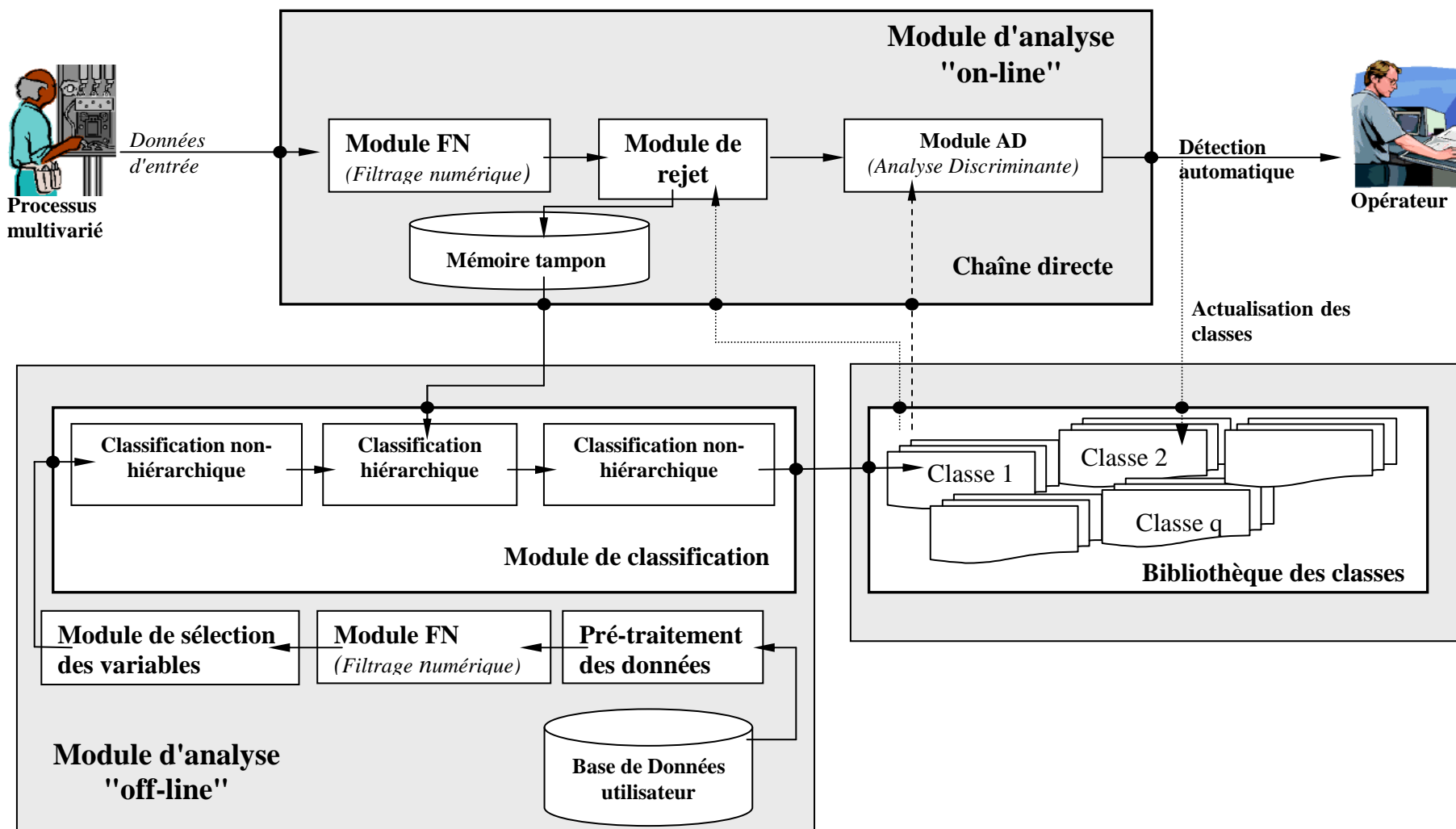


figure III.4 – Méthode FNAD

3.4.2 Types de causes assignables étudiées

Selon les lois de la physique aucun système ne peut fonctionner isolé dans la nature. Les processus de fabrication ou les systèmes industriels n'échappent pas à ce principe général. Par la suite, un système industriel peut être défini comme un ensemble, susceptible d'évoluer dans le temps, composé d'éléments interagissant entre eux et avec l'extérieur, et qui sont réunis pour assurer une mission (Dubuisson [27]).

Tout processus de fabrication est soumis à différentes sources de variation. C'est pour cette raison que dans une production en série deux pièces ne sont jamais parfaitement identiques et on trouvera toujours de petites différences d'une pièce à l'autre. Nous allons dans ce qui suit analyser ces différentes sources de variation qu'on peut grouper en deux catégories: causes communes de variation et causes assignables de variation.

3.4.2.1 Causes communes de variation

Certaines variations sont inhérentes au processus et portent le nom de "*causes communes*". Elles sont dues au hasard qui reste présent plus ou moins dans chaque processus de fabrication. Du point de vue des physiciens les changements imprévisibles d'une quantité variant dans le temps sont connus sous le nom de bruits. Souvent, ces causes communes portent le nom de bruit blanc. Le bruit blanc est un signal complètement non-corrélé et avec une densité spectrale constante dans toute la bande des fréquences. Le bruit blanc joue un rôle important dans la modélisation des phénomènes aléatoires. En général, on fait l'hypothèse que les bruits sur lesquels on ne dispose pas d'information particulière sont des bruits blancs centrés stationnaires et gaussiens. Dans la majorité des cas, le modèle gaussien est satisfaisant pour décrire les variations inhérentes au processus, donc pour modéliser les bruits. Il y a aussi des situations pratiques où l'utilisation de la loi normale pour modéliser les causes communes n'est pas adéquate. Gunter [44], Polansky et al. [96], Chen et Pearn [11], Castagliola [9], Somerville et Montgomery [111] ont analysé les conséquences de la non-normalité sur le calcul des limites de contrôle pour les cartes de contrôle et sur l'exactitude des coefficients de capacité pour le processus. Souvay et Magnien [112] donnent l'exemple des défauts de symétrie qui suivent une loi demi-normale et des défauts de type coaxialité qui suivent une loi de Rayleigh.

L'aléatoire fait l'objet de la théorie statistique et, par conséquent, on sait très bien modéliser ces causes communes et prévoir ainsi le comportement d'un processus qui est régi uniquement par ce types de sources de variation. L'ensemble de ces causes communes forme la variabilité intrinsèque du processus. Ces types de variations sont très difficiles à éliminer sans restructurer d'une manière radicale le système ce qui coûte généralement très cher. On peut, dans la majorité des cas, envisager de les réduire en adoptant une série de mesures adéquates mais il faut se résigner à l'idée qu'elles ne vont jamais disparaître complètement et qu'on doit "vivre avec". Un processus qui est affecté seulement par des causes communes de variation est dit "sous-contrôle" ou en "fonctionnement normal". Par "fonctionnement normal" du processus on comprend son déroulement dans un état qui assure la réalisation d'un objectif prédéfini. Le fonctionnement est considéré anormal dans le cas contraire. On pourra citer comme exemple des causes communes le jeu dans la chaîne cinématique de la machine, le défaut de la broche de la machine, etc...

3.4.2.2 Causes assignables de variation

Une deuxième catégorie est constituée par les variations qui apparaissent accidentellement dans le processus et qui portent le nom de "causes assignables". Ce type de variations est plus rare mais son effet est beaucoup plus dérangeant que l'effet des causes communes. Les causes assignables doivent être identifiées et éliminées afin d'améliorer la qualité du processus. Les cartes de contrôle aident à mettre en évidence l'apparition de ces causes assignables et par conséquent en les éliminant elles contribuent à la réduction de la variabilité du processus. Un processus qui est affecté par les causes assignables est dit "hors-contrôle". On pourra citer comme exemples de causes assignables le dérèglement d'un outil, l'usure d'un outil, une mauvaise lubrification, l'opérateur, la mauvaise qualité d'un matériel utilisé, etc...

Dans notre cas, nous avons étudié deux types de causes assignables intervenant souvent dans un processus de fabrication: le saut en échelon (figure III.5.a) et la dérive en rampe (figure III.5.b).

Les causes assignables de type saut en échelon surgissent instantanément dans le processus et consistent dans le passage de la caractéristique étudiée du niveau μ_0 au niveau μ_1 . Elles sont modélisées par l'équation suivante:

$$x_k = \begin{cases} \mu_0, & k = 0, \dots, t_0 \\ \mu_1 = \mu_0 + \text{Amp}, & k = t_0 + 1, \dots \end{cases} \quad (3.20)$$

où "Amp" représente l'amplitude du dérèglement.

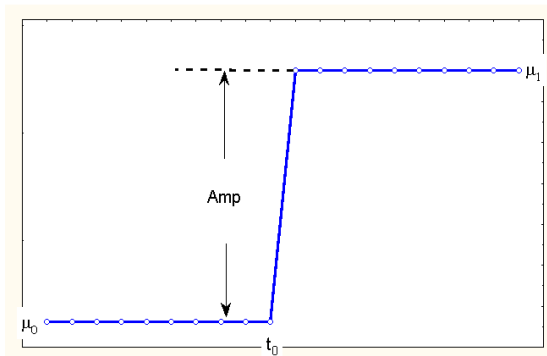


figure III.5.a – saut en échelon

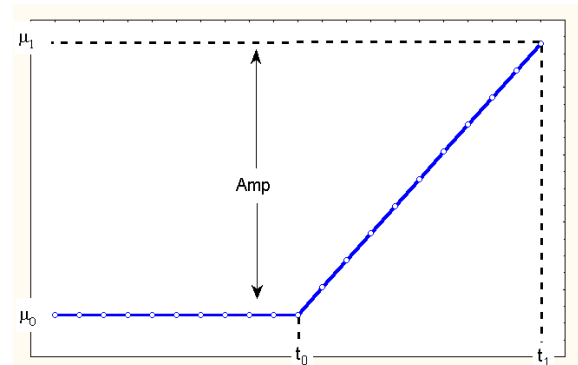


figure III.5.b – dérive en rampe

On peut citer comme exemples de causes assignables de type échelon: le changement des matières premières, le changement d'opérateur, etc...

La dérive en rampe est une cause assignable qui se manifeste par une augmentation progressive de la caractéristique analysée. Ce type de cause assignable est modélisé par l'équation suivante:

$$x(t) = \begin{cases} \mu_0, & t < t_0 \\ \mu_1 = \mu_0 + \frac{t-t_0}{t_1-t_0} \times \text{Amp}, & t_0 < t \leq t_1 \end{cases} \quad (3.21)$$

où "Amp" représente l'amplitude de la dérive.

On peut citer comme exemples de causes assignables de type dérive en rampe: l'usure de l'outillage, etc...

3.4.2.3 Causes communes et causes assignables de variation

A la sortie du processus, les causes communes (bruits) et les causes assignables (signaux) s'additionnent comme le montre la figure III.6.a et la figure III.6.b. Nous avons simulé dans ce cas un bruit blanc de moyenne nulle et d'écart-type $\sigma = 1$ superposé à un saut en échelon et une dérive en rampe d'amplitude égale à 4σ .

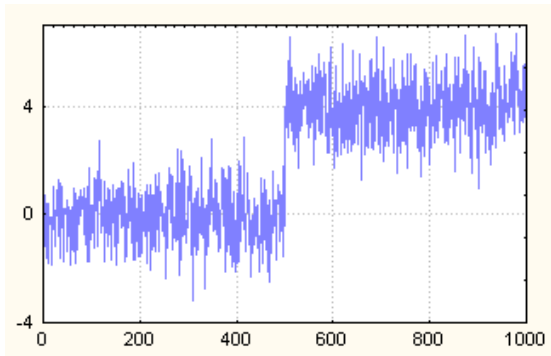


figure III.6.a – saut en échelon + bruit

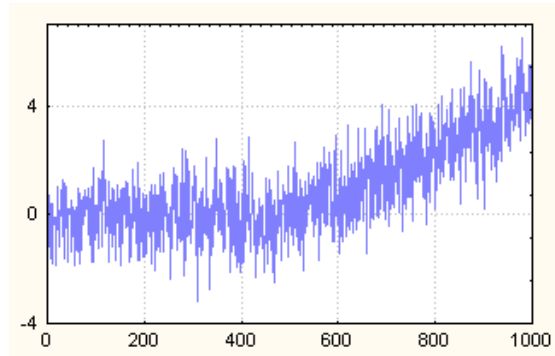


figure III.6.b – dérive en rampe + bruit

3.4.3 Analyse spectrale des causes assignables

Dans le précédent paragraphe nous avons modélisé dans le temps deux causes spéciales: le saut en échelon et la dérive en rampe. Dans ce paragraphe nous allons analyser leur composition spectrale. Cette analyse spectrale est nécessaire pour la définition de la méthode de filtrage numérique et nous l'utiliserons également pour le choix du type de filtre.

L'analyse spectrale du saut en échelon est faite en fonction de plusieurs paramètres (figure III.7.a):

- ✓ le nombre total d'échantillons: N
- ✓ le nombre d'échantillons correspondant au saut: Ns
- ✓ l'amplitude du saut: Amp

Dans la figure III.7.b on peut constater que le spectre du saut en échelon est distribué selon une fonction sinus cardinal. Si Ns et Amp sont constants, l'amplitude du spectre diminue quand N augmente. Pour N et Ns constants, l'amplitude du spectre augmente au fur et à mesure que l'amplitude du saut augmente (figure III.7.c).

Finalement, quand N et Amp sont constantes, l'amplitude du spectre augmente quand N_s augmente (figure III.7.d).

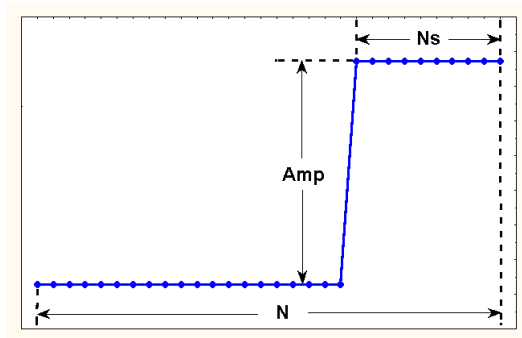


figure III.7.a - saut en échelon

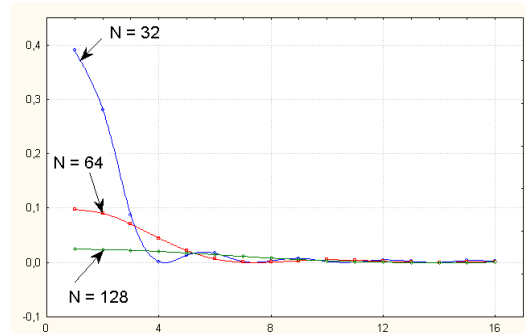


figure III.7.b - spectre en fonction de N

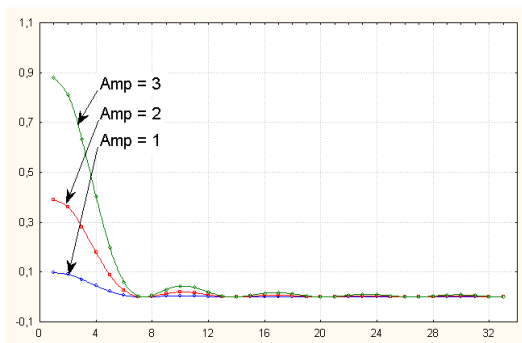


figure III.7.c - spectre en fonction de Amp

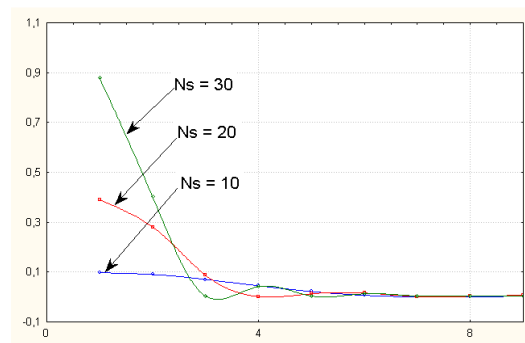


figure III.7.d - spectre en fonction de N_s

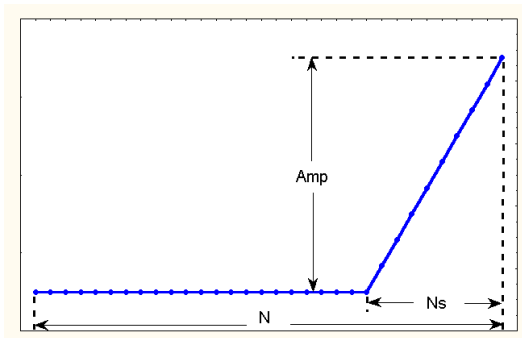


figure III.8.a - dérive en rampe

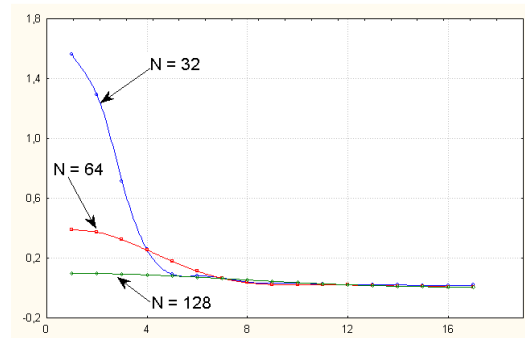


figure III.8.b - spectre en fonction de N

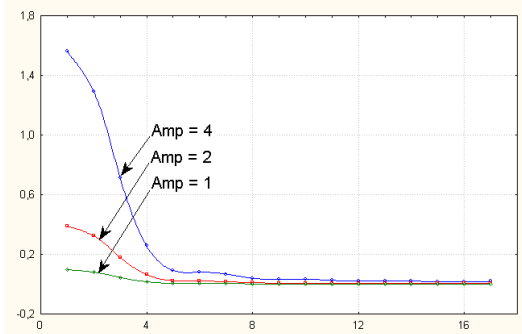


figure III.8.c - spectre en fonction de Amp

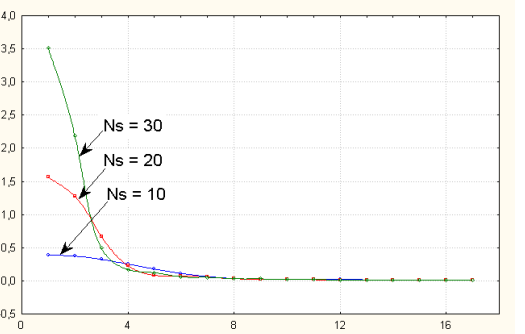


figure III.8.d - spectre en fonction de N_s

L'analyse spectrale de la dérive en rampe est faite en fonction des mêmes paramètres (figure III.8.a). Le spectre de la dérive en rampe est distribué selon une fonction qui

ressemble à un sinus cardinal au carré. En observant les figures III.8.b, III.8.c et III.8.d nous aboutissons aux mêmes conclusions que pour le saut en échelon. L'amplitude du spectre augmente proportionnellement à l'amplitude du saut (Amp) et avec le nombre des échantillons correspondant au saut (Ns). Elle diminue quand le nombre total d'échantillons (N) augmente.

Après avoir étudié la composition spectrale du saut en échelon et de la dérive en rampe en fonction de plusieurs paramètres, on peut conclure que leur présence se manifeste, principalement, à des basses-fréquences. Le bruit blanc, par contre, a des composantes spectrales uniformément distribuées dans toute la bande des fréquences (figures III.9.d et III.10.d). Par conséquent, quand les deux causes de variation sont présentes simultanément les composantes spectrales du bruit s'additionnent aux composantes spectrales de la cause assignables. Le spectre qui en résulte est décrit dans la figure III.9.f pour le saut en échelon et dans la figure III.10.f pour la dérive en rampe (nous avons simulé dans les deux cas pour 50 observations un bruit blanc ayant une distribution normale de moyenne nulle et écart-type $\sigma = 1$). Un dérèglement (saut en échelon; dérive en rampe), d'amplitude 1σ , a été simulé à partir de l'échantillon 25 (figure III.9.b et figure III.10.b).

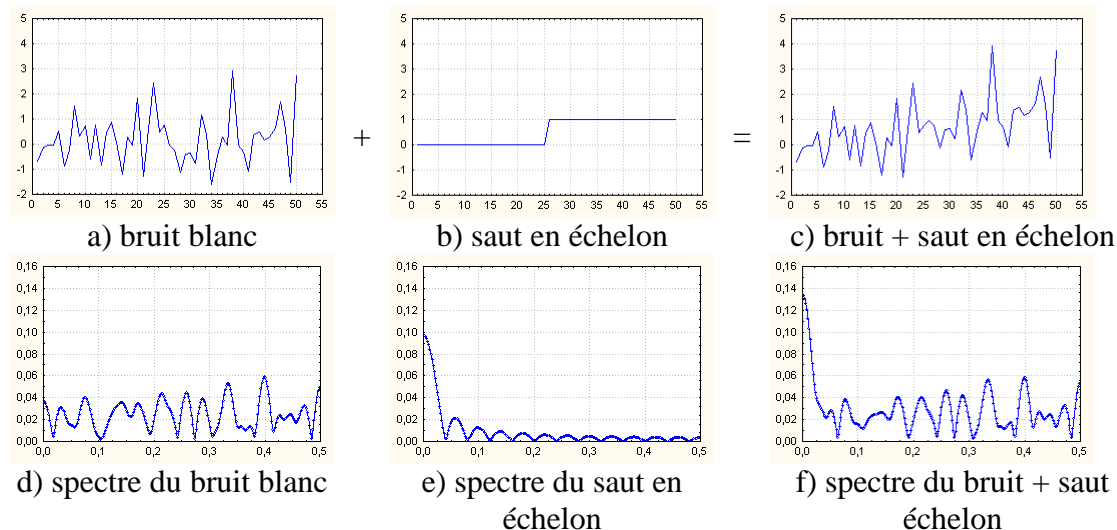


figure III.9 - évolution du saut en échelon dans le temps et en fréquence

On peut remarquer en comparant les figures III.9.d et III.9.f et les figures III.10.d et III.10.f, que les deux causes assignables se manifestent par des composantes spectrales aux basses-fréquences. Aux hautes-fréquences le spectre du bruit reste presque inchangé.

Nous avons également analysé le spectre de deux causes assignables en présence d'un bruit qui n'est pas distribué selon une distribution normale. Nous avons simulé des bruits ayant des distributions non-normales (uniforme, demi-normale et Rayleigh). Les premières 500 observations parmi les 1000 observations simulées correspondent au fonctionnement normal du procédé c'est-à-dire à la présence du bruit uniquement. A partir de l'observation 501 nous avons simulé des saut en échelon et des dérèglements de type dérive en rampe de différentes amplitudes à partir de 0.5σ et jusqu'à 4σ . Nous présentons ci-dessus les résultats obtenus pour les différentes configurations de bruit.

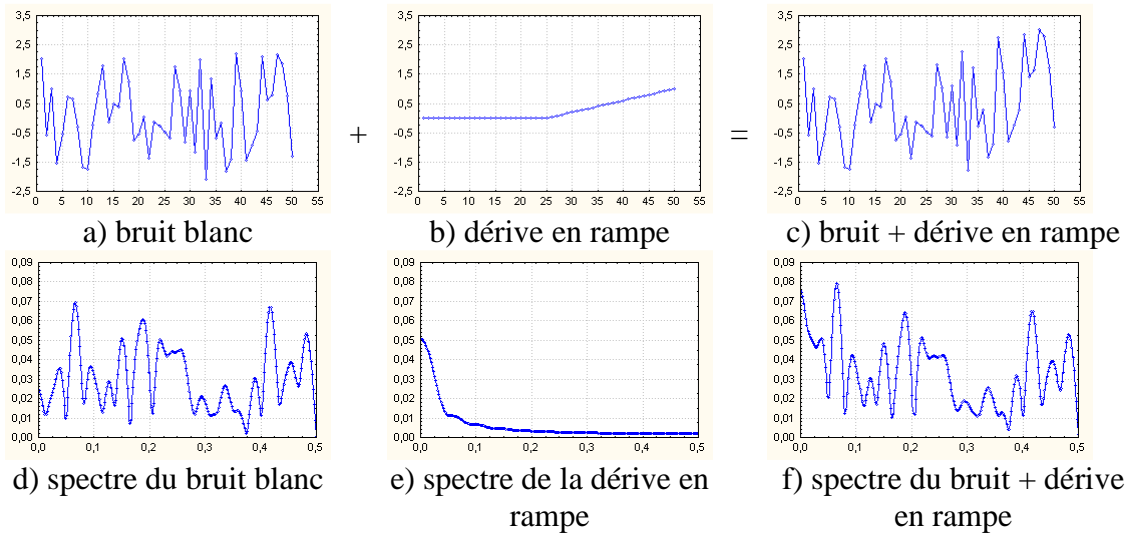


figure III.10 - évolution de la dérive en rampe dans le temps et en fréquence

Bruit de distribution uniforme

Nous avons simulé un bruit uniforme entre $[-\sqrt{3}$ et $\sqrt{3}]$ c'est-à-dire de moyenne nulle et écart-type $\sigma = 1$ pour conserver la similitude avec le bruit blanc gaussien. Nous présentons dans la figure III.10.1 le spectre du bruit et des saut en échelon dans la présence du bruit. On peut facilement remarquer qu la présence d'une cause spéciale de type saut en échelon est mise en évidence par des composantes spectrales situées à des basses fréquences. Les mêmes conclusions restent valables pour les causes assignables de type dérive en rampe et par conséquent nous avons préféré de ne le pas représenter ici.

Bruit de distribution demi-normale

Ce type de bruit caractérise les défauts de symétrie car la grandeur mesurée est prise en valeur absolue. Dans la présence d'un dérèglement l'allure de la distribution change et s'approche d'autant plus de la distribution normale que l'amplitude du dérèglement est grande. Dans la figure III.10.2 et dans la figure III.10.3 nous présentons les courbes de densité spectrale pour les deux types de dérèglages étudiés en présence d'un bruit ayant une distribution demi-normale. Ayant en vue que le bruit demi-normal n'a pas une moyenne nulle mais égale à $m_{dn} = 0,798\sigma_n$, où σ_n est l'écart-type du bruit gaussien, nous avons construit les courbes de densité spectrale dans les figures 7 et 8 après avoir soustrait la moyenne du bruit.

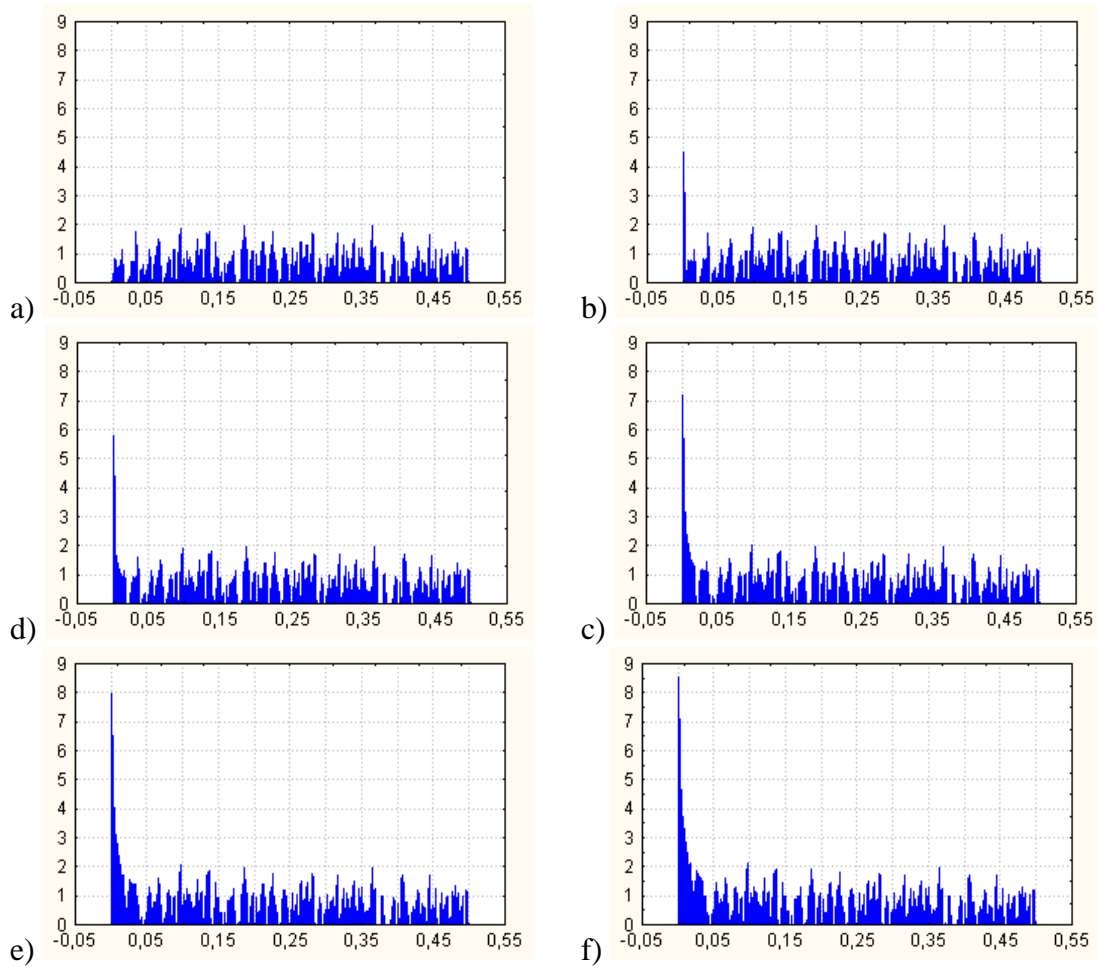


figure III.10.1 (échelle logarithmique) – densité spectrale du bruit uniforme (a) et des sauts en échelon bruités d'amplitude 0.5σ (b), 1σ (c), 2σ (d), 3σ (e) et 4σ (f)

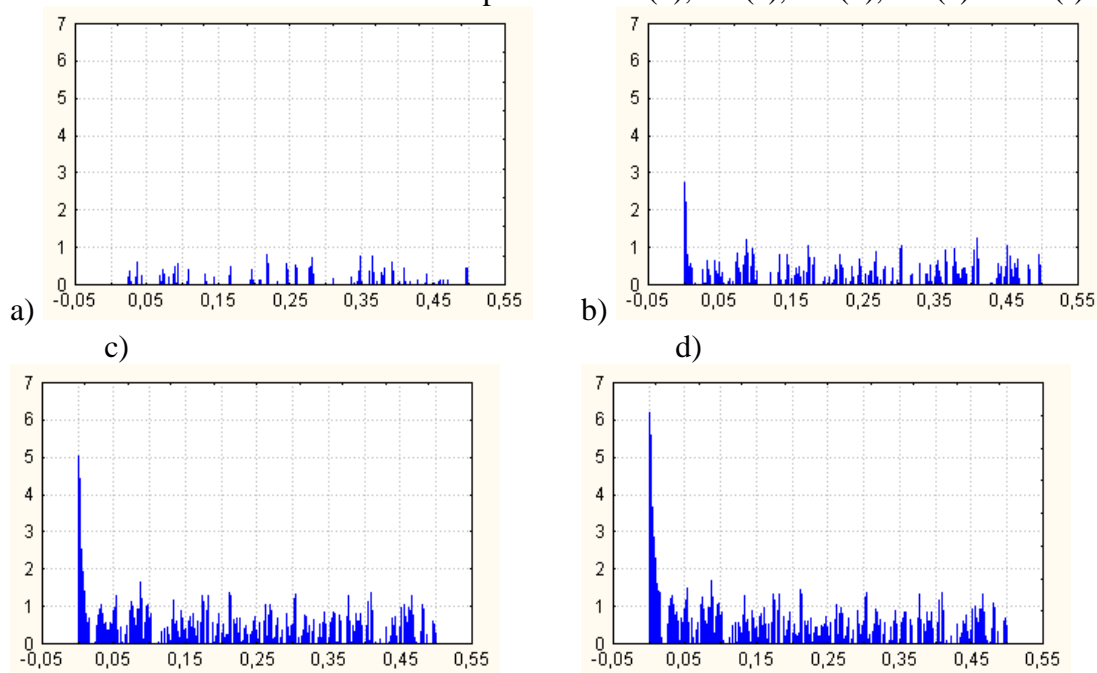


figure III.10.2 (échelle logarithmique) – densité spectrale du bruit de distribution demi-normale (a) et des sauts en échelon bruités d'amplitude 1σ (b), 2σ (c), 3σ (d)

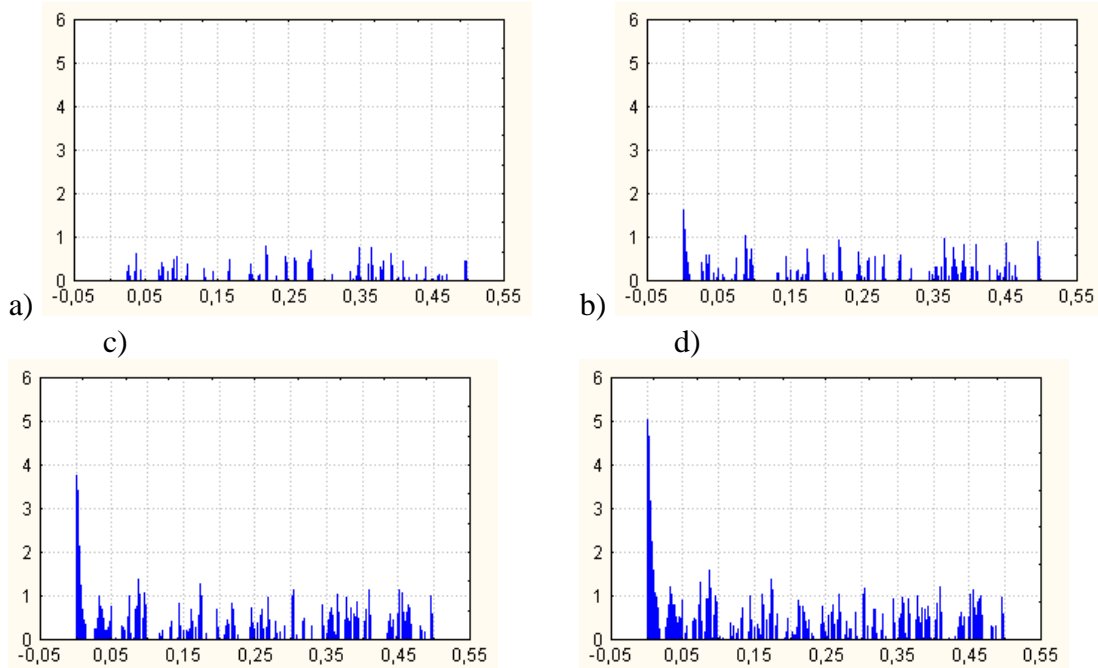


figure III.10.3 (échelle logarithmique) – densité spectrale du bruit de distribution demi-normale (a) et des dérivés en rampe bruités d'amplitude 1σ (b), 2σ (c), 3σ (d)

Nous pouvons remarquer (figures III.10.2 et III.10.3) que les deux causes assignables sont présentes au basses fréquences. Entre autre, par rapport aux situations antérieures (bruit blanc et bruit uniforme), la détection est plus facile car on peut identifier dans toute la bande spectrale la présence des composantes qui n'existaient pas avant l'apparition de la cause assignable. L'explication pour ce modification de spectre réside dans les changements qui se produisent dans la distribution du bruit suite à l'apparition de la cause spéciale. Dans la figure III.10.4 nous pouvons constater comment la distribution demi-normale du bruit change en fonction de l'apparition d'une cause spéciale de type saut échelon ou dérive en rampe.

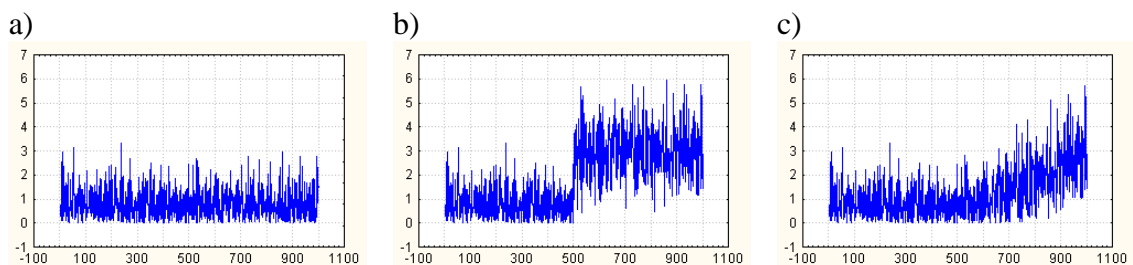


figure III.10.4 - bruit demi-normale (a); bruit demi-normale et saut en échelon d'amplitude 3σ (b); bruit demi-normale et dérive en rampe d'amplitude 3σ (c)

Bruit de distribution Rayleigh

Ce type de bruit caractérise entre autre les défauts de symétrie. Tout comme pour le bruit demi-normale, dans la présence d'un dérèglement l'allure de la distribution change et s'approche d'autant plus de la distribution normale que l'amplitude du

dérégage est grande. Tenant compte que la composition spectrale du bruit de type Rayleigh n'est pas très différente de celle du bruit de distribution demi-normale, dans la figure III.10.5 on donne seulement le spectre du bruit, du saut en échelon et de la dérive en rampe d'amplitude 4. Sachant que la moyenne du bruit de type Rayleigh est $m_r = 1,25\sigma_n$, où σ_n est l'écart-type du bruit gaussien nous avons soustrait cette moyenne avant de construire les représentations spectrales de la figure III.10.5. Les conclusions qu'on a énoncé pour le bruit de distribution demi-normale restent valables pour le bruit de distribution Rayleigh aussi.

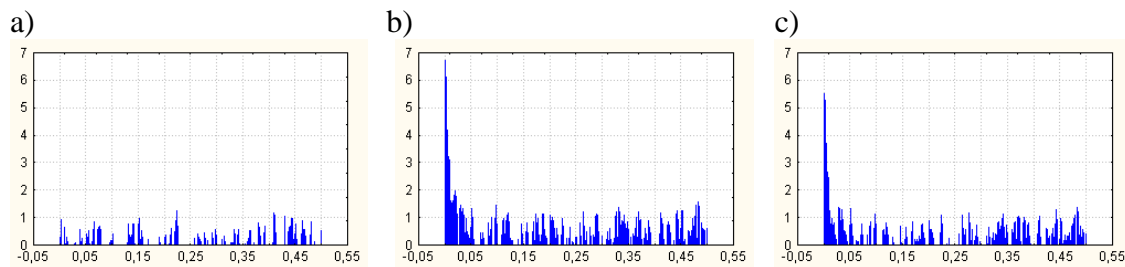


figure III.10.5 - bruit de type Rayleigh (a); bruit de type Rayleigh et saut échelon d'amplitude 4σ (b); bruit de type Rayleigh et dérive en rampe d'amplitude 4σ (c)

Selon les résultats obtenus, il en résulte que la présence des causes assignables de type saut en échelon ou dérive en rampe est identifiable à des basses fréquences dans le domaine spectral et cela indifféremment de la nature du bruit. En plus, dans certaines situations, bruit de distribution demi-normale ou Rayleigh (caractérisant, par exemple, les défauts de symétrie et respectivement de coaxialité), la détection des dérégages est encore plus facile car des changements sont observables dans tout le spectre du signal. Cela vient du changement de la distribution du bruit dans la présence du dérégage et, donc, au rapprochement de la distribution normale au fur et à mesure que l'amplitude du dérégage augmente.

3.4.4 Module d'analyse "off-line"

Dans le module d'analyse "off-line" toute l'information disponible dans la base de données doit être extraite afin de concevoir une bibliothèque de classes. Cette étape étant très importante, elle doit être effectuée avec beaucoup d'attention et de patience même si elle est très coûteuse en temps et volume de calcul. A l'issue de ce module on dispose d'un espace de représentation de dimension \mathbf{p} et d'un ensemble de \mathbf{q} classes, chacune correspondant à un mode de fonctionnement du système.

3.4.4.1 Base de données utilisateur

Les informations contenues dans la base de données peuvent être issues d'un historique de l'évolution du processus ou d'expérimentations. L'historique correspond à l'enregistrement de nombreuses variables de processus (ou des variables de qualité du produit) sur une longue période. On obtient ainsi des grandes quantités de données dans lesquelles seuls quelques états de fonctionnement sont

présents, voir même uniquement le mode de fonctionnement normal. Dans ce cas, l'information sur ce(s) mode(s) de fonctionnement est riche.

Dans le cas d'expérimentations, les données sont beaucoup moins nombreuses mais le nombre de modes de fonctionnement est généralement plus élevé que dans le cas précédent. On est, donc, en présence d'un ensemble de données qui décrit mieux les différents modes de fonctionnement mais l'information disponible sur chaque mode de fonctionnement est moins riche. Dans ce cas, il est nécessaire de mettre au point un protocole d'expériences comparable à celui utilisé en modélisation des systèmes.

Le système d'information correspond à un ensemble de moyens et de processus assurant le traitement de l'information pour aider à la prise de décision. Plus précisément un système d'information est constitué d'une base de données et d'un ensemble de processus capables de traiter ces données. La rapidité et l'efficacité de la prise de décision sont directement liées à la qualité du système d'information. Or, nous sommes convaincus, qu'aujourd'hui, suite à l'extraordinaire évolution de l'électronique et de l'informatique, les industriels ont la possibilité d'enregistrer dans des bases de données une quantité impressionnante d'informations liées au fonctionnement du processus. Ces données sont, parfois, stockées même si leur utilisation ultérieure n'est pas très claire et précise. Le recueil des données doit, pourtant, être fait avec le plus grand soin car de la qualité et de la représentativité de ces données dépend l'analyse ultérieure du processus.

3.4.4.2 Module de pré-traitement des données

Comme nous l'avons précisé dans la partie introductive, ce module vise à rendre la base de données exploitable pour les futurs traitements. Il sert pour la résolution des problèmes suivants:

- ✓ transformation des données;
- ✓ prise en compte des observations douteuses (les observation incomplètes ou aberrantes);

Le premier aspect, est lié au fait qu'en pratique, pour les variables quantitatives, il existe des problèmes d'homogénéité d'unités de mesure. L'utilisation brute des variables dans les calculs de distances entre les individus donnera de façon implicite plus de poids aux variables de plus forte dispersion, annulant presque complètement l'effet des autres variables (voir l'exemple de la distance de Mahalanobis traité dans le chapitre 1). En normalisant les données on peut résoudre ce problème mais cette opération présente parfois des inconvénients majeurs, Kresta et al. [67]. En effet, la normalisation peut modifier la nature des relations établies entre les variables et donc faire disparaître la facilité de l'interprétation physique du phénomène. Kresta et al. [67] recommandent l'utilisation du même facteur d'échelle quand on veut conserver la relation qui existe entre les variables (dans le cas où celles-ci ont la même unité de mesure).

Le deuxième aspect concerne le fait que les données obtenues sont susceptibles de contenir deux types d'observations indésirables pour la suite des traitements: *les valeurs aberrantes* et *les observations manquantes*.

Les valeurs aberrantes introduisent un biais pour l'opération de centrage et réduction des données, ou encore dans les métriques utilisées par les différentes méthodes statistiques multidimensionnelles (comme l'analyse discriminante, les méthodes de classification, etc...). En plus, forcer un point aberrant à appartenir à une classe "déforme" cette classe. Par conséquent, ayant en vue tous ces inconvénients, plusieurs recherches ont été réalisées afin d'améliorer la détection des valeurs aberrantes.

Kobi [63] propose une méthodologie de détection en ligne des valeurs aberrantes basée sur l'analyse en composantes principales. La détection se fait, soit par une inspection visuelle des projections dans les plans factoriels, soit par le calcul d'une distance entre les différentes observations. Dans un premier temps, la contribution de l'observation aberrante à la construction des axes factoriels est calculée et, dans un deuxième temps, la corrélation entre les axes factoriels et les variables du départ est déterminée. De cette manière la variable en défaut est identifiée aussi, et ceci est un avantage majeur dans le cas multivarié. L'inconvénient majeur de la méthode est qu'il n'y a aucune distance statistique qui soit satisfaisante pour la détection de la valeur aberrante et donc, l'approche reste plutôt visuelle. Rao [100], Gnanadesikan et Kettenring [36] ont suggéré l'utilisation de la somme des carrés des dernières composantes principales pour la définition de la distance. Une amélioration a été proposée par Hawkins [48] en donnant le même poids à toutes les composantes et obtenant ainsi une distance similaire à celle de Mahalanobis. Gnanadesikan et Kettering [36] ont proposé une distance qui se calcule en multipliant la composante principale par sa variance. Cette distance permet la mise en évidence des observations qui ont un grand effet sur le premier axe factoriel. Hawkins [49] propose une distance qui est basée sur le calcul de la valeur maximale des dernières composantes principales normées. Kobi [63] montre que l'efficacité de ces distances est fonction du nombre de composantes principales retenues et qu'il n'y a pas une règle générale pour fixer ce nombre. Il y a également la possibilité de tester individuellement chaque variable pour détecter la présence d'éventuelles valeurs aberrantes. Plusieurs tests univariés sont connus dans la littérature et nous ne rappelons ici que quelques uns des plus utilisés. Grubbs [42, 43] propose deux tests pour la détection des valeurs aberrantes: le premier est basé sur l'écart des observations par rapport à la moyenne; le deuxième est construit sur le rapport des variances de l'échantillon total et l'échantillon réduit obtenu après avoir éliminé l'observation suspectée d'être aberrante. Dixon [22] propose une fonction discriminante permettant de résoudre le problème de la détection des valeurs aberrantes. Malheureusement cette fonction discriminante prend plusieurs formes suivant le nombre de mesures et suivant qu'on teste les valeurs les plus faibles ou les plus fortes. Dixon [23], Grubbs [43] et Barnett et Lewis [5] donnent plusieurs variantes du test de Dixon. Celles-ci sont obtenues en combinant les différentes configurations que l'on désire prendre en compte. Cochran [16] propose un test basé sur le rapport entre la valeur extrême et la moyenne de l'échantillon. Il est surtout utilisé lorsque les mesures suivent une autre loi de distribution que la loi normale. Kobi [63] a réalisé une analyse et une comparaison de ces différents tests pour la détection en ligne d'observations aberrantes.

Etant donné que les équations des filtres numériques nous ramènent aux modèles des séries temporelles AR (Auto Regressive), MA (Moving Average) et ARMA (Auto Regressive and Moving Average), les méthodes permettant la détection des valeurs aberrantes dans ces situations présentent un intérêt tout particulier dans le cadre de la méthode FNAD. De nombreuses recherches existent dans cette direction et nous allons

mentionner par la suite seulement quelques unes des plus représentatives. Fox [32] a été parmi les premiers à s'intéresser à ce type de problème. Il a proposé une méthode basée sur le rapport de vraisemblance afin de tester deux hypothèses liées à l'absence et à la présence des valeurs aberrantes. Muirhead [87] utilise le logarithme de la fonction de vraisemblance dans un modèle auto-régressif mixte. Bruce et Martin [8] et Ledolter [71] étendent la technique de la suppression d'une observation à un modèle ARIMA (Auto Regressive Integrate Moving Average) pour mesurer l'influence d'un ensemble d'observations sur l'estimation des paramètres et l'estimation de la variance de la séquence d'innovation. Abraham et Chuang [1] adaptent une statistique, développée par Draper et John [26] et utilisée dans un modèle linéaire, à des séries temporelles. Une étude comparative des différents tests pour la détection des valeurs aberrantes dans une série temporelle a été réalisée par Kobi [63].

Les observations manquantes posent des problèmes pour les méthodes classiques d'analyse car on obtient une matrice où il manque des éléments; l'élimination systématique de l'observation ou de la variable incomplète n'est pas une solution toujours satisfaisante car elle peut nous priver d'une information précieuse. La question d'une meilleure utilisation de ce type d'observations se pose alors. Trois solutions sont possibles pour remédier à cette situation: adapter les calculs pour une prise en compte intelligente des données manquantes; attribuer des valeurs aux données manquantes (reconstitution des données manquantes) ce qui permet ensuite l'utilisation des techniques statistiques classiques; supprimer les individus qui présentent des données manquantes (faute de mieux).

Une combinaison des approches ci-dessus mentionnées est attrayante pour les raisons suivantes: d'une part, si une observation ou une variable a un grand nombre de mesures manquantes, la perte d'information résultant de son exclusion n'est pas significative alors que son utilisation dans le calcul produira probablement un biais significatif (dans ce cas l'élimination est un meilleur choix); d'autre part, si une observation ou une variable a peu de valeurs manquantes, alors la perte d'information est significative si l'approche par élimination est employée (dans ce cas, l'adaptation est le meilleur choix).

3.4.4.3 Module de Filtrage Numérique

Nous allons d'abord analyser dans ce paragraphe la nécessité de l'utilisation du filtrage numérique dans le schéma représenté dans la figure III.4. Ensuite, les paramètres caractérisant qualitativement un bon filtre numérique seront exposés. Finalement, nous allons justifier le choix du filtre et nous allons déterminer ses paramètres optimaux.

3.4.4.3.1 Pourquoi le filtrage numérique est nécessaire?

Nous avons représenté sur le schéma de la méthode FNAD (figure III.4), deux modules de filtrage numérique. Le premier module est placé avant le module de classification et le deuxième module est placé dans la chaîne directe avant le module d'analyse discriminante (AD).

Pourquoi le filtrage numérique est nécessaire dans ce schéma de la méthode FNAD ? Pour répondre à cette question, il faut penser à l'objectif de la méthode FNAD. Nous voulons réaliser un diagnostic automatique des différentes causes assignables

préalablement détectées sur le processus. Or, très souvent, en pratique, ces causes assignables sont de très faible amplitude. Dans ce cas, si on veut créer une bibliothèque de classes ou effectuer directement une analyse discriminante, le taux d'erreur de classification ou de classement est très élevé.

Evidemment, le taux d'erreur dépend d'un grand nombre de facteurs comme le nombre de variables contrôlées, le nombre d'observations disponibles sur le processus, le nombre d'observations disponibles pour chaque classe de dérèglement, le nombre et la nature des dérèglements présents, etc... Pour cette raison, généralement, l'évaluation du taux d'erreur reste difficile à réaliser.

Néanmoins, pour se faire une idée de la valeur que peut prendre le taux d'erreur dans une opération de classification, nous avons simulé une situation très simple: pour un nombre de **p** variables (compris entre 3 et 10) ayant une distribution normale de moyenne $\mu = 0$ et écart-type $\sigma = 1$, nous avons simulé des sauts en échelon d'amplitude comprise entre **Amp** = 1σ et **Amp** = 4σ ; nous avons ensuite effectué une classification par des nuées dynamiques (d'une manière générale, la méthode des nuées dynamiques donne exactement le nombre de classes souhaitées aussi différentes entre elles que possible; pour une discussion plus détaillée sur les méthodes de classification voir le paragraphe 3.4.4.5.1).

Le taux moyen d'erreur pour cette méthode de classification a été calculé sur 200 simulations. Les résultats obtenus sont présentés dans le tableau III.1.

On peut remarquer que le taux d'erreur augmente au fur et à mesure que le nombre de variables augmente. Ainsi, par exemple, quand l'amplitude du saut est égale à 4σ , nous avons obtenu, pour trois variables, un taux d'erreur égal à 7,81 % et un taux d'erreur égal à 30,24 % quand le nombre de variables est égal à dix.

tableau III.1

	Amp = 1	Amp = 2	Amp = 3	Amp = 4
p = 3	61,61 %	44,4 %	20,32 %	7,81 %
p = 4	67,34 %	48,66 %	25,34 %	9,72 %
p = 5	71,78 %	55,43%	29,99 %	13,23 %
p = 10	93,25 %	70,82 %	52,52 %	30,24 %

On constate, également, que le taux d'erreur augmente quand l'amplitude du saut diminue, ce qui était d'ailleurs prévisible. Ainsi, pour un nombre de trois variables le taux d'erreur passe de 7,81% pour un dérèglement d'amplitude 4σ à 61,61 % pour un dérèglement d'amplitude 1σ .

On peut conclure, d'après ces résultats, qu'il est très difficile de bien classer les observations quand l'amplitude du dérèglement est faible. En regardant le tableau III.1, on peut voir que l'opération de classification par des nuées dynamiques donne des résultats complètement aberrants dans le cas où l'amplitude du dérèglement est égale à 1σ et le nombre de variables supérieur à dix. Même pour un nombre plus réduit de variables et pour des amplitudes supérieures à 1σ le taux d'erreur de classification reste élevé. Or, la conception de la méthode FNAD repose sur l'utilisation d'une bibliothèque des classes. Dans cette bibliothèque, une classe correspond à un type de dérèglement enregistré pendant le fonctionnement du processus. Elle est décrite par un vecteur des moyennes et une matrice de variance-covariance. Si la composition des classes est mal déterminée suite à de nombreuses erreurs de classification, il est clair que les résultats obtenus à la sortie du module d'analyse discriminante seront faussés. Ceci est dû au fait que l'analyse discriminante utilise l'information contenue dans la bibliothèque de classes afin d'attribuer des nouvelles observations à une des classes existantes.

Il est, donc, impérativement nécessaire de réduire le taux d'erreur pour que la méthode reste fiable et crédible. La réduction du taux d'erreur reste une tâche difficile sans réduire la variabilité (la variation due au bruit) pour mettre plus clairement en évidence la cause assignable, c'est-à-dire sans avoir filtré au préalable les données pour réduire leur variabilité.

3.4.4.3.2 Les paramètres caractéristiques du filtrage numérique

Pour caractériser les filtres numériques dans le temps, il faut analyser leur réponse au signal échelon appliqué à l'entrée. Trois paramètres sont fréquemment utilisés pour comparer les filtres dans le domaine temporel:

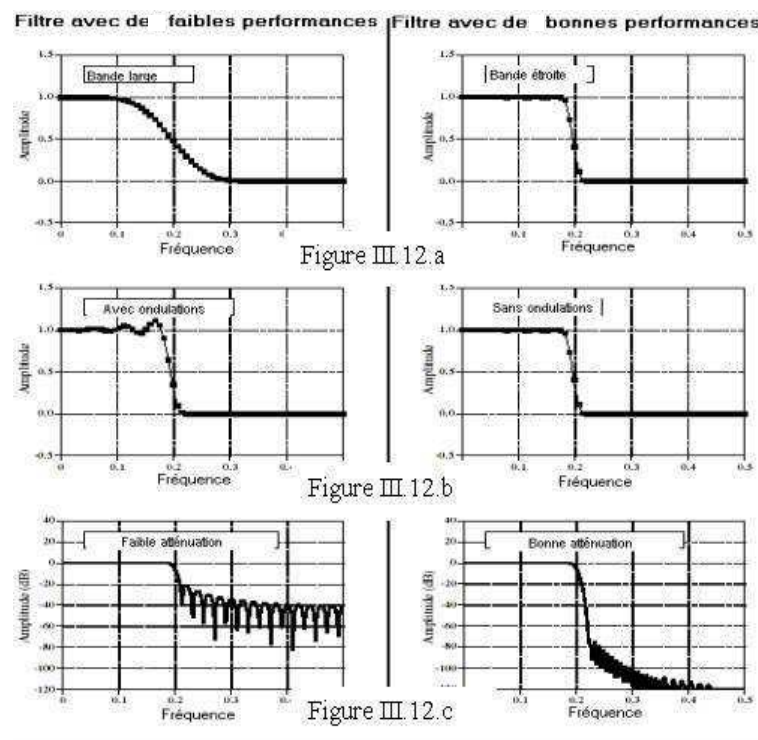
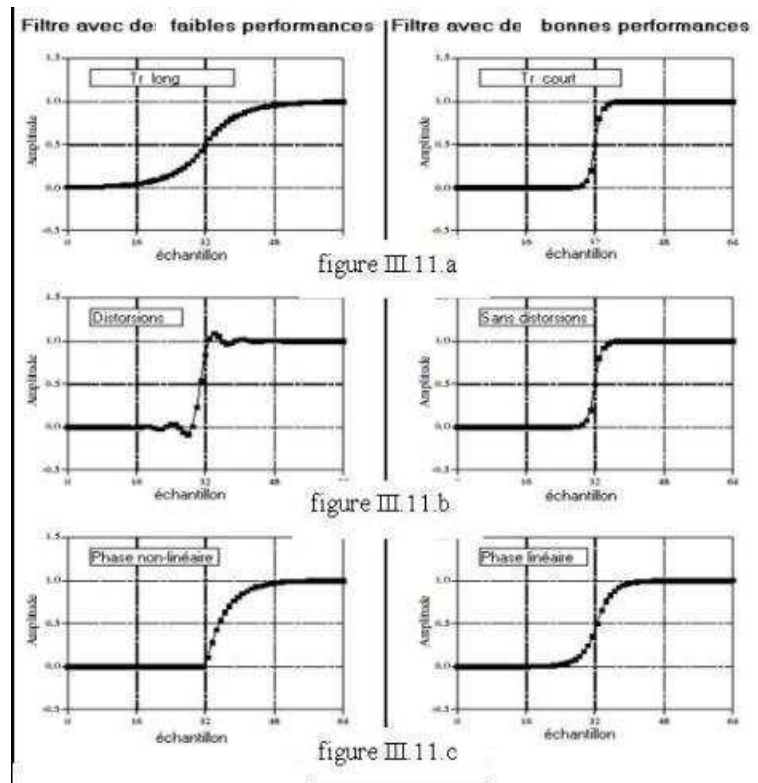
- ✓ le temps de réponse: T_r ;
- ✓ la distorsion: D ;
- ✓ la phase: Ph ;

Un bon filtre numérique doit avoir un **temps de réponse** court (figure III.11.a). En pratique, on considère que c'est le nombre d'échantillons compris entre 10% et 90% de l'amplitude du saut en échelon qui définit le temps de réponse du filtre. La **distorsion** (figure III.11.b), se manifeste par le changement de l'amplitude des échantillons dans le signal temporel. Un bon filtre numérique ne doit pas induire des distorsions dans le signal. Finalement, il est souvent souhaitable que la moitié supérieure de la fonction de réponse au signal en échelon du filtre soit symétrique avec la moitié inférieure, comme le montre la figure III.11.c. Cette symétrie est nécessaire pour que les montées et les descentes dans le signal aient le même aspect et elle est nommée **phase linéaire**. Un bon filtre numérique est caractérisé par une phase linéaire.

Pour caractériser les filtres numériques en fréquence, il faut analyser leur fonction de transfert (ou leur gabarit). Trois paramètres sont fréquemment utilisés pour comparer les filtres dans le domaine fréquentiel:

- ✓ la largeur de la bande de transition: BT ;
- ✓ les ondulations dans la bande passante: O ;
- ✓ l'atténuation dans la bande d'arrêt: AT ;

Pour bien discriminer entre des fréquences voisines, un bon filtre numérique doit avoir une bande de transition étroite (figure III.12.a). Pour que les fréquences situées dans la bande passante puissent traverser le filtre sans être altérées il faut que la bande passante soit sans ondulations. Un bon filtre numérique présente une bande passante plate (figure III.12.b). Finalement, pour bien bloquer le passage des fréquences situées dans la bande d'arrêt du filtre, il faut que la bande d'arrêt ait une bonne atténuation (figure III.12.c).



3.4.4.3 Choix du filtre numérique

Nous avons constaté (paragraphe 3.4.3) que la présence des causes assignables de type saut en échelon et dérive en rampe se manifeste, principalement, à des basses-fréquences. Le bruit, par contre, a des composantes spectrales uniformément distribuées dans toute la bande spectrale.

Nous voulons réduire la variation du signal et mettre plus clairement en évidence la présence de la cause assignable. Cet objectif peut être atteint si on utilise des filtres numériques de type passe-bas. Les filtres passe-bas laissent passer uniquement les basses fréquences et bloquent toutes les autres fréquences du signal. Ainsi, par le filtrage, on arrive à augmenter la proportion des composantes spectrales de la cause assignable par rapport aux composantes spectrales du bruit. Finalement, la variation caractérisant le bruit est réduite et la présence de la cause assignable est plus nette. Par la méthode de filtrage passe-bas que l'on propose, le bruit n'est jamais complètement éliminé, il est seulement réduit, car une grande partie de ses composantes spectrales sont éliminées.

Dans notre cas, l'idéal serait d'avoir un filtre numérique de type passe-bas ayant les caractéristiques suivantes:

- ✓ **temps de réponse court:**
Il est souhaitable d'avoir un temps de réponse le plus court possible car on veut détecter la cause assignable au plus vite dans le domaine du temps. Ce paramètre est analogue à la POM_1 d'une carte de contrôle;
- ✓ **sans distorsions:**
Il faut que le signal passe à travers le filtre sans subir des modifications de forme ou d'aspect. Cette propriété est encore plus importante dans le cadre de la méthode FNAD parce que chaque type de cause assignable doit être identifiée selon ses caractéristiques;
- ✓ **bande de transition étroite:**
Pour mieux séparer les bandes des fréquences il est nécessaire d'avoir un filtre avec une bande de transition étroite;
- ✓ **sans ondulations dans la bande passante:**
Il faut que la forme de la cause assignable reste la même après avoir éliminé le bruit. Pour cela, toutes les fréquences la caractérisant doivent traverser le filtre avec le même gain, c'est-à-dire que la bande passante doit être plate, sans ondulations;
- ✓ **bonne atténuation dans la bande d'arrêt:**
Afin de réduire au maximum l'amplitude des composantes spectrales situées dans la bande d'arrêt, celle-ci doit présenter une bonne atténuation;
- ✓ **caractéristique de phase linéaire:**
C'est une caractéristique de filtrage qui relève, dans notre cas, plutôt de l'esthétique que de la pratique. Pour mieux comprendre cet aspect, prenons l'exemple d'un signal de forme rectangulaire qui passe par le filtre. Si, à la sortie du filtre, la forme du front montant est similaire à la forme du front descendant on dit que le filtre a une caractéristique de phase linéaire. S'il y a une différence entre la forme du front montant et la forme du front descendant alors on dit que le filtre a une caractéristique de phase non-linéaire;

Le filtre idéal, ayant toutes les caractéristiques ci-dessus mentionnées, n'existe pas. En pratique, nous devons nous contenter d'approcher cette situation idéale. Pour cela, il faut faire des compromis.

Nous commençons notre analyse par une étude des caractéristiques fréquentielles pour différents filtres numériques passe-bas de type RII (Réponse Impulsionnelle Infinie) et de type RIF (Réponse Impulsionnelle Finie). Dans la figure III.13.a, nous présentons quelques fonctions de transfert des filtres passe-bas de type Butterworth ayant une fréquence de coupure $FC=0.1$ Hz. On peut remarquer que la fonction de transfert du filtre s'approche de la fonction de transfert idéale passe-bas quand l'ordre du filtre (L) augmente. La bande passante et la bande d'arrêt sont sans oscillations et la largeur de la bande de transition est d'autant plus petite que l'ordre du filtre est élevé. On peut remarquer, également, que la bande passante est de gain unitaire et que la bande d'arrêt présente une bonne atténuation.

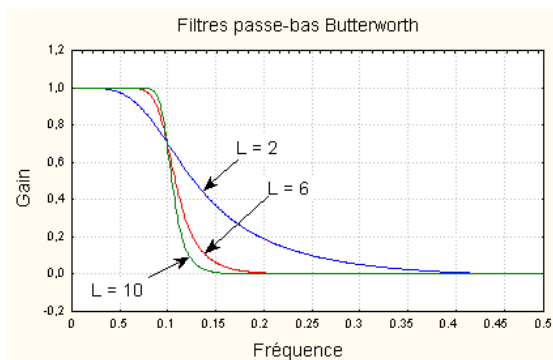


figure III.13.a

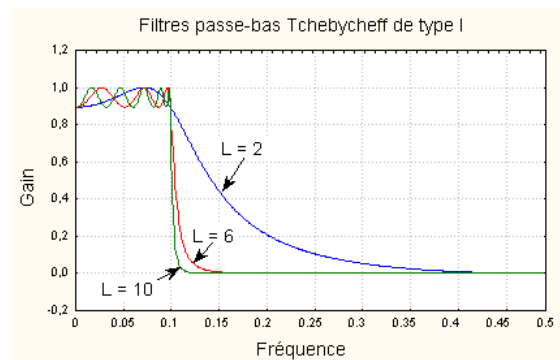


figure III.13.b

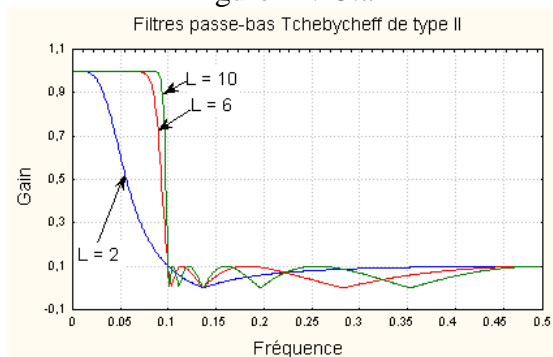


figure III.13.c

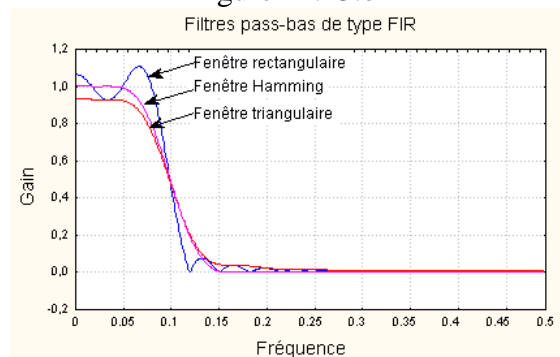


figure III.13.d

Dans la figure III.13.b nous présentons les fonctions de transfert de filtres passe-bas Tchebycheff de type I, ayant une fréquence de coupure $FC=0.1$ Hz. Ce type de filtre est caractérisé par des oscillations dans la bande passante. Par contre, la bande d'arrêt présente une très bonne atténuation et elle est sans oscillations. On remarque que la largeur de la bande de transition est d'autant plus petite que l'ordre du filtre (L) est élevé. L'avantage de ce type de filtre, par rapport au filtre Butterworth, est que pour le même ordre L , la bande de transition est plus étroite. Cet avantage est obtenu au détriment des oscillations dans la bande passante. Si on réduit l'amplitude des oscillations dans la bande passante, on augmente, implicitement, la largeur de la bande de transition.

Dans la figure III.13.c nous présentons les fonctions de transfert de filtres passe-bas Tchebycheff de type II, ayant une fréquence de coupure $FC=0.1$ Hz. Par rapport au filtre Tchebycheff de type I, ce type de filtre se caractérise par des oscillations dans la bande d'arrêt et par une bande passante sans oscillations. La bande de transition est d'autant

plus étroite que l'ordre du filtre augmente. Tout comme le filtre Tchebycheff de type I, le filtre Tchebycheff de type II a une bande de transition plus étroite que le filtre Butterworth, pour le même ordre du filtre. L'inconvénient de ce type de filtre est la présence d'oscillations dans la bande d'arrêt.

Les filtres Butterworth et Tchebycheff sont des filtres de type RII. Dans la figure III.13.d, nous présentons la caractéristique de transfert d'un filtre passe-bas de type RIF, d'ordre $L=30$ et ayant une fréquence de coupure $FC=0.1$ Hz. Pour que la réponse impulsionnelle soit finie, ces types de filtres ont une fonction de transfert qui s'obtient par une convolution entre la fonction de transfert passe-bas idéale et une fenêtre de pondération. Il existe plusieurs types de fenêtres de pondération. Dans la figure III.13.d, nous en avons utilisé trois: la fenêtre rectangulaire, la fenêtre triangulaire et la fenêtre Hamming. On peut remarquer que, pour une fenêtre de pondération rectangulaire, on a des oscillations tant dans la bande passante que dans la bande d'arrêt. Par l'utilisation d'une fenêtre triangulaire ces oscillations sont plus réduites mais on constate que le gain dans la bande passante est inférieur à 1 et l'atténuation dans la bande d'arrêt n'est pas très bonne non plus. Une amélioration est obtenue en utilisant une fenêtre de pondération de type Hamming. Pour que la bande de transition soit étroite, l'ordre du filtre doit être très élevé. On peut constater, en comparant les figures III.13.d et III.13.a, que pour obtenir une fonction de transfert similaire, en fréquence, un filtre FIR doit avoir un ordre très élevé (dans ce cas $L=30$), alors qu'un filtre Butterworth nécessite un ordre beaucoup plus réduit (dans ce cas $L=6$).

Après avoir constaté ce qui se passe en fréquence, il est nécessaire de poursuivre l'analyse dans le temps. Dans les figures III.14.a-d, nous présentons la réponse des précédents filtres au saut en échelon, d'amplitude égale à 1, simulé à partir de l'observation 20.

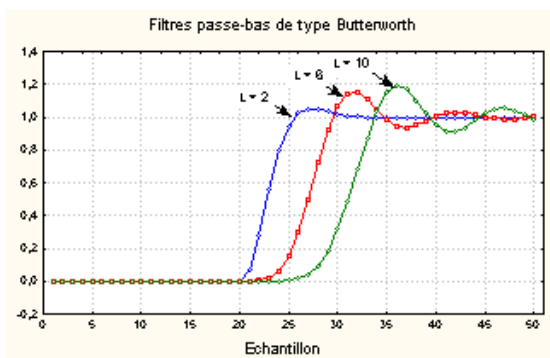


figure III.14.a

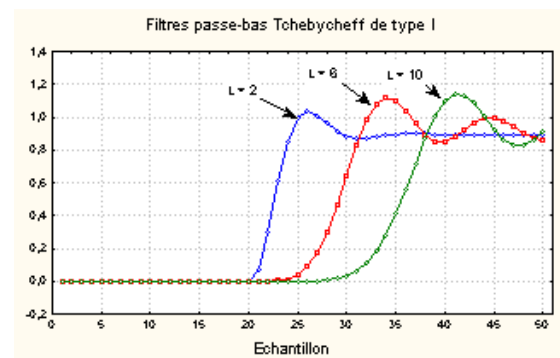


figure III.14.b

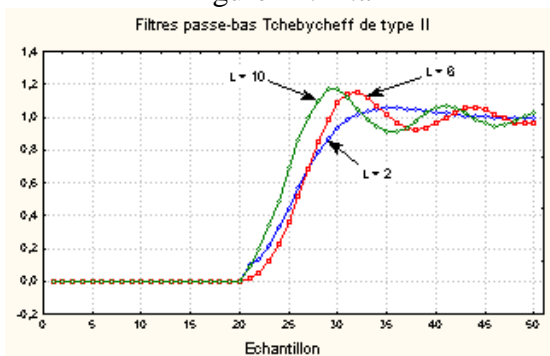


figure III.14.c

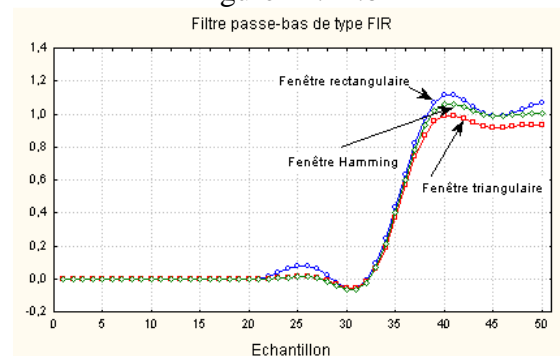


figure III.14.d

On peut remarquer (figure III.14.a) que le temps de réponse du filtre passe-bas Butterworth augmente au fur et à mesure que l'ordre du filtre augmente. En plus, quand l'ordre du filtre L est élevé, il y a des distorsions du signal. Les distorsions, qui se manifestent sous la forme d'une courbe sinusoïdale amortie, ont une amplitude d'autant plus élevée que l'ordre du filtre est grand.

Le temps de réponse du filtre passe-bas Tchebycheff de type I ayant l'ordre $L=2$ est légèrement inférieur à celui du filtre Butterworth de même ordre. Le problème est que le signal est atténué à la sortie du filtre. Au fur et à mesure que l'ordre du filtre augmente le temps de réponse augmente également et on peut remarquer, en comparant les figures III.14.a et III.14.b, que le filtre Tchebycheff de type I est plus lent que celui de Butterworth. La distorsion du signal est d'autant plus importante que l'ordre du filtre est grand.

Pour le filtre passe-bas Tchebycheff de type II (figure III.14.c) on constate que, au fur et à mesure que l'ordre du filtre augmente, le temps de réponse se réduit et la distorsion du signal augmente. Par rapport au filtre Tchebycheff de type I, le temps de réponse du filtre Tchebycheff de type II est nettement plus réduit. On voit aussi, en comparant les figures III.14.a et III.14.c, que les filtres Butterworth d'ordre $L=2$ et Tchebycheff d'ordre $L=10$ ont approximativement le même temps de réponse. L'avantage du filtre Butterworth par rapport au filtre Tchebycheff est que la distorsion du signal, dans le temps, est plus réduite. Par contre, le filtre Tchebycheff donne des meilleurs résultats en fréquence.

Finalement, on peut constater, en regardant la figure III.14.d, que les filtres FIR ont un temps de réponse beaucoup plus élevé que les filtres RII ayant les mêmes performances en fréquence. Si on veut améliorer leur temps de réponse on doit réduire leur ordre mais ainsi, leurs performances en fréquence sont très faibles.

Pour conclure, après avoir analysé les deux caractéristiques majeures (en temps et en fréquence) des filtres numériques, notre choix se porte sur les filtres passe-bas de type Butterworth. Nous avons choisi l'utilisation des filtres Butterworth pour les raisons suivantes:

- ✓ ils ne présentent pas des ondulations, ni dans la bande passante ni dans la bande d'arrêt.
- ✓ ils se caractérisent par une bonne atténuation dans la bande d'arrêt.
- ✓ ils ont une caractéristique de phase linéaire.
- ✓ ils ont un temps de réponse plus court par rapport aux autres filtres présentés; le temps de réponse est d'autant plus court que l'ordre du filtre est réduit.
- ✓ les distorsions du signal sont d'autant plus réduites que l'ordre du filtre est diminué; on remarque sur la figure III.14.a que pour un filtre d'ordre $L=2$, le signal ne contient presque pas des distorsions.

Certes, les filtres Butterworth ont une bande de transition plus large que les filtres Tchebycheff, mais cela ne représente pas un très grand inconvénient par rapport aux autres critères déjà mentionnés.

3.4.4.3.4 Choix des paramètres du filtre passe-bas de type Butterworth

Après avoir choisi l'utilisation du filtre passe-bas de type Butterworth pour le filtrage numérique, on veut déterminer dans ce sous-paragraphe les valeurs les plus adéquates pour l'ordre (L) et la fréquence de coupure (FC) du filtre.

Si on regarde la figure III.14.a, on constate que le temps de réponse du filtre augmente quand son ordre augmente. Or, nous voulons un filtre qui répond le plus vite possible à tout changement de l'entrée. Par conséquent, il va falloir réduire au maximum l'ordre du filtre. Le problème est que, de cette façon, on obtient un filtre avec de faibles caractéristiques en fréquence (figure III.13.a). Cela veut dire que, au fur et à mesure que l'on réduit l'ordre du filtre, la bande de transition devient plus large et par conséquent la discrimination des fréquences est moins bonne. Nous avons vu, dans le paragraphe 3.4.3 que la présence de la cause assignable (saut en échelon ou dérive en rampe) se manifeste principalement à des basses fréquences. Si l'ordre du filtre est réduit, beaucoup de composantes spectrales propres au bruit restent encore dans le signal, suite à l'augmentation de la largeur de la bande de transition. Par conséquent, la variabilité du signal à la sortie du filtre reste élevée et il sera difficile de mettre en évidence la présence de la cause assignable. Pour remédier à cet inconvénient, on peut envisager la réduction de la fréquence de coupure (FC) du filtre (figure III.15.a).

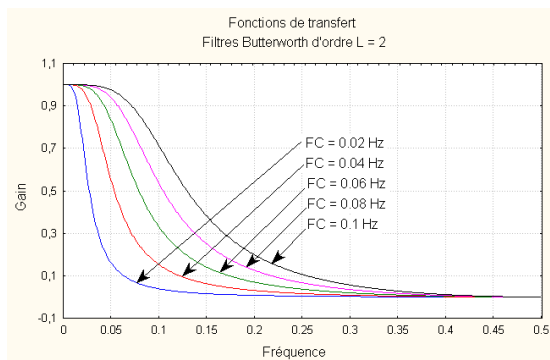


figure III.15.a

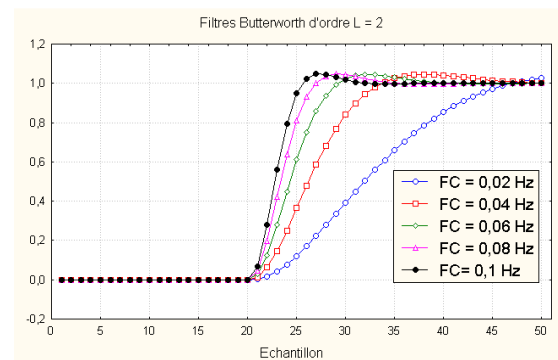


figure III.15.b

Dans la figure III.15.a, on peut constater que, pour le même ordre du filtre $L = 2$, au fur et à mesure que l'on réduit la fréquence de coupure (FC), la largeur de la bande d'arrêt augmente et en même temps la largeur de la bande passante et de la bande de transition diminuent. Cela veut dire que, seules les composantes spectrales de très basse fréquence peuvent passer par le filtre. Or, nous avons vu que ces composantes sont en majorité celles qui caractérisent la cause assignable. Ainsi, le bruit est réduit, et la présence de la cause assignable sera plus clairement mise en évidence. Le problème est que de cette manière on augmente aussi le temps de réponse du filtre (figure III.15.b).

On se trouve, donc, confronté à un problème d'optimisation. Il faut trouver le bon compromis entre l'ordre (L) et la fréquence de coupure (FC) du filtre. La question qui se pose maintenant est: quelle critère choisir pour résoudre ce problème d'optimisation? Pour répondre à cette question, on doit penser à l'utilité du filtrage numérique. Or, nous avons constaté (paragraphe 3.4.4.3.1) que le filtrage numérique est nécessaire afin de réduire le nombre d'erreurs de classification et le nombre d'erreurs de classement.

Nous allons donc chercher ce compromis pour minimiser le nombre moyen d'erreurs de classement de la méthode d'analyse discriminante et en même temps pour réduire le temps de réponse du filtre passe-bas de type Butterworth.

Pour cela, nous allons d'abord donner une définition du temps de réponse du filtre sensiblement différente de la définition classique du temps de réponse utilisée dans le filtrage numérique. On rappelle que la définition classique du temps de réponse

est: le nombre d'échantillons dans la réponse du filtre compris entre 10% et 90% de l'amplitude du saut échelon. Nous allons définir le temps de réponse (TR) du filtre comme étant le nombre d'échantillons compris entre le moment d'apparition du saut échelon (E_1) et jusqu'à ce que la sortie du filtre atteigne 90% de l'amplitude du saut (E_k), voir la figure III.16. Par conséquent: $TR = \text{nombre d'échantillons entre l'échantillon } (E_k) \text{ et l'échantillon } (E_1)$.

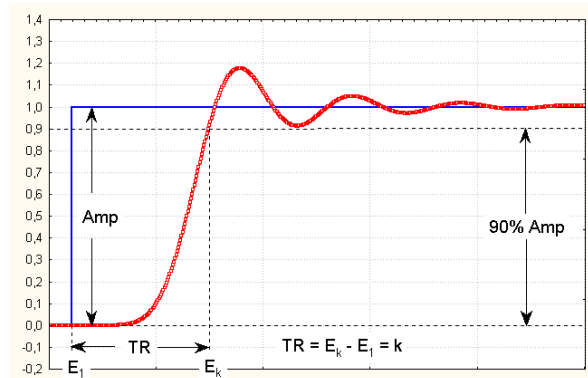


figure III.16 - Définition du temps de réponse pour le filtre numérique

Nous avons calculé le temps de réponse des filtres passe-bas Butterworth en fonction de l'ordre et de la fréquence de coupure du filtre pour un saut échelon d'amplitude égale à 1. Les résultats sont présentés dans le tableau III.2. Nous avons choisi d'étudier le temps de réponse des filtres passe-bas Butterworth ayant l'ordre compris entre 2 et 20 et la fréquence de coupure comprise entre 0.01 Hz et 0.1 Hz. Dans la figure III.17 nous donnons sous forme graphique le temps de réponse, présenté dans le tableau III.2, en fonction de l'ordre et la fréquence de coupure du filtre.

tableau III.2

L \ FC	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
2	42	21	14	11	8	7	6	5	5	4
4	63	32	21	16	13	10	9	8	7	6
6	85	43	28	21	17	14	12	11	9	8
8	107	53	36	27	21	18	15	13	12	11
10	129	64	43	32	26	21	18	16	14	13
12	150	75	50	37	30	25	21	19	16	15
14	172	86	57	43	34	28	24	21	20	17
16	193	96	64	48	38	32	27	24	21	19
18	224	107	71	53	43	35	30	26	23	21
20	236	118	78	59	47	39	33	29	26	23

On peut remarquer que pour un certain ordre fixé (L) du filtre, le temps de réponse augmente au fur et à mesure que la fréquence de coupure (FC) du filtre est réduite. Ainsi, par exemple, pour un filtre passe-bas de type Butterworth d'ordre $L = 6$, le temps de réponse (TR) est égal à 8 échantillons quand la fréquence de coupure est $FC = 0.1$ Hz, alors que pour le même filtre, le temps de réponse est $TR = 85$ quand la fréquence de coupure est $FC = 0.01$ Hz (voir le tableau III.2).

On constate, également, que pour une certaine fréquence de coupure fixée, le temps de réponse est d'autant plus élevé que l'ordre du filtre est grand. Ainsi, pour une fréquence de coupure $FC = 0.04$ Hz, par exemple, on a un temps de réponse $TR = 11$ échantillons pour un filtre passe-bas de type Butterworth d'ordre $L = 2$ et un temps de réponse $TR = 59$ pour un filtre d'ordre $L = 20$.

Nous avons également vérifié que les temps de réponse dans le tableau III.2 restent les mêmes pour des amplitudes de saut variant entre 1.5σ et 3σ .

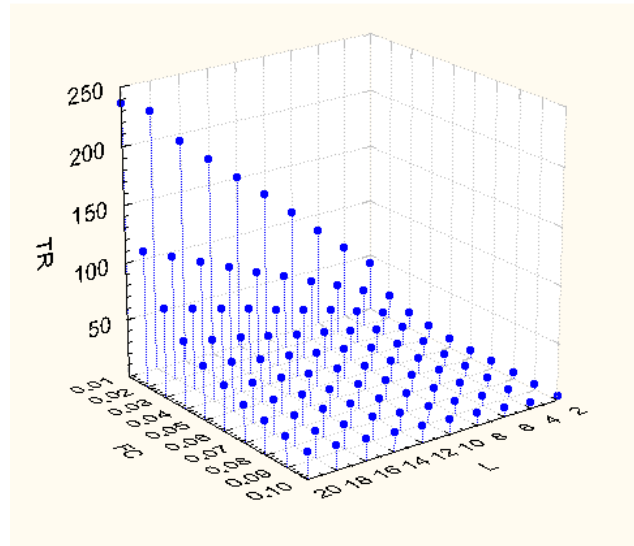


figure III.17 - Temps de réponse (TR) fonction de la fréquence de coupure (FC) et de l'ordre (L) du filtre

Si l'on considérait seulement le temps de réponse dans le critère d'optimisation, on aurait pu affirmer que pour le minimiser il faudrait choisir un filtre passe-bas de type Butterworth ayant l'ordre le plus petit possible et la fréquence de coupure la plus grande possible. Mais nous devons tenir compte, également, pour optimiser notre critère, du pourcentage d'erreur (%ERR) de classement de la méthode d'analyse discriminante. Par conséquent, nous allons étudier l'influence de l'ordre (L) et de la fréquence de coupure (FC) du filtre sur le pourcentage moyen d'erreur (%ERR) de classement, et ensuite nous essayerons d'optimiser simultanément le pourcentage d'erreur de classement et le temps de réponse du filtre.

Comme le nombre d'observations des différentes classes de fonctionnement du processus est limité, les règles de décision diffèrent des règles optimales ce qui entraînent une augmentation de l'erreur de classification. Jain et Chandrasekaran [60] préconisent que le nombre minimum d'observations par classes soit de cinq à dix fois le nombre de variables. De nombreuses recherches ont été effectuées pour trouver la relation entre l'erreur de classification, la taille des classes, le nombre de variables et la complexité du classifieur (Fukunaga [34]; Raudys et Jain [102]). Fukunaga [34] montre clairement qu'un multiple fixe du nombre des variables ne peut pas être utilisé pour déterminer la taille de l'ensemble d'apprentissage. Raudys et Jain [102] proposent d'utiliser l'estimation de l'erreur de classification pour déterminer la taille de l'ensemble d'apprentissage (pour atteindre un niveau fixé de précision d'apprentissage) et de test (pour estimer l'erreur de classification avec une certaine précision). La taille minimale de l'ensemble d'apprentissage dépend de la règle de décision choisie, du nombre de variables, de la probabilité d'erreur de classification et de la précision d'apprentissage

souhaitée. Afin d'estimer la taille requise de l'ensemble d'apprentissage, les auteurs recommandent d'utiliser l'estimation non paramétrique de l'augmentation de l'erreur de classification. Si l'estimation non paramétrique de l'augmentation de l'erreur de classification est petite en comparaison de l'estimée de la probabilité d'erreur nous pouvons alors conclure que la taille de l'ensemble d'apprentissage est suffisante.

La taille minimale de l'ensemble de test peut être déterminée à partir de la variance de l'estimée de l'erreur de classification. Pour cela, il est nécessaire de fixer une valeur ou un intervalle de la vraie probabilité d'erreur. La taille obtenue est alors la taille nécessaire pour que l'estimée de la probabilité d'erreur ne dévie pas de la valeur vraie plus que d'un pourcentage fixé.

Dans le cas du diagnostic, le problème se complique davantage, car il faut déterminer les seuils de rejet d'ambiguïté et de distance qui influencent l'erreur de classification.

Etant donné qu'il est très difficile d'obtenir une expression analytique de la probabilité d'erreur pour un problème de classification particulier, la probabilité d'erreur doit être estimée à partir des échantillons disponibles. Plusieurs estimateurs de l'erreur de classification ont été proposés (Raudys et Jain [102]). Lorsque seulement un nombre fini d'observations est disponible le pourcentage d'observations mal classées (Fukunaga [34]) est un estimateur non-biaisé de l'erreur de classement.

En partant de cette constatation, nous avons effectué des simulations, en prenant en compte différentes configurations des filtres passe-bas Butterworth, pour déterminer le taux d'erreur de classification moyen à la sortie du module d'analyse discriminante. Nous avons choisi, pour des considérations pratiques, uniquement les configurations des filtres présentant un temps de réponse inférieur ou égal à 21 observations (nous considérons que les filtres ayant un temps de réponse supérieur à cette valeur sont trop "lents" pour être utilisés dans une procédure efficace de détection).

Pour un nombre de variables compris entre trois et cinq, nous avons généré des dérèglages de type saut en échelon d'amplitude 1σ (où σ est l'écart-type des variables) pour chaque variable. Le nombre d'observations correspondant à chaque classe de dérèglement est égal à 100. Le nombre total d'observation est en fonction du nombre de variables: 700 pour trois variables; 900 pour quatre variables; 1100 pour cinq variables.

Le nombre de classes de fonctionnement varie également en fonction du nombre de variables. Ainsi pour trois variables on a quatre classes (trois classes de dérèglement et la classe de fonctionnement normal du processus), pour quatre variables on a cinq classes et finalement pour cinq variables on a six classes de fonctionnement.

Dans le tableau III.3 nous présentons, pour 200 simulations, les résultats obtenus (le taux d'erreur de classification moyen) pour des filtres passe-bas d'ordres $L = 2$ et $L = 4$ et ayant différentes fréquences de coupure (FC). Les figures III.18 et III.19 traduisent sous forme graphique ces différentes informations. La première ligne du tableau donne le taux d'erreur de classification moyen pour des données "brutes", c'est-à-dire sans l'utilisation du filtrage numérique. La dernière colonne du tableau III.3 donne les temps de réponses (TR) des différents filtres. On peut remarquer que même un filtrage avec une fréquence de coupure relativement élevée ($FC=0.1$ Hz) réduit d'une manière significative (approximativement 50%) le taux d'erreur de classification.

tableau III.3 – le taux d'erreur de classification (en %)

L	FC	p = 3	p = 4	p = 5	TR
-	-	57,08	65,01	69,69	0
2	0.02	3.51	4.37	5.59	21
	0.03	6.20	8.09	9.63	14
	0.04	9.78	12.28	14.96	11
	0.05	13.17	16.57	20.22	8
	0.06	16.66	20.61	24.75	7
	0.07	19.51	24.56	28.85	6
	0.08	22.45	27.90	32.38	5
	0.09	25.08	30.59	35.58	5
4	0.03	5.38	6.94	8.60	21
	0.04	8.78	10.79	13.27	16
	0.05	12.44	15.18	18.34	13
	0.06	15.46	18.90	22.88	10
	0.07	18.31	22.78	27.17	9
	0.08	21.14	26.28	30.82	8
	0.09	23.63	29.31	33.97	7
	0.1	26.04	31.86	36.94	6

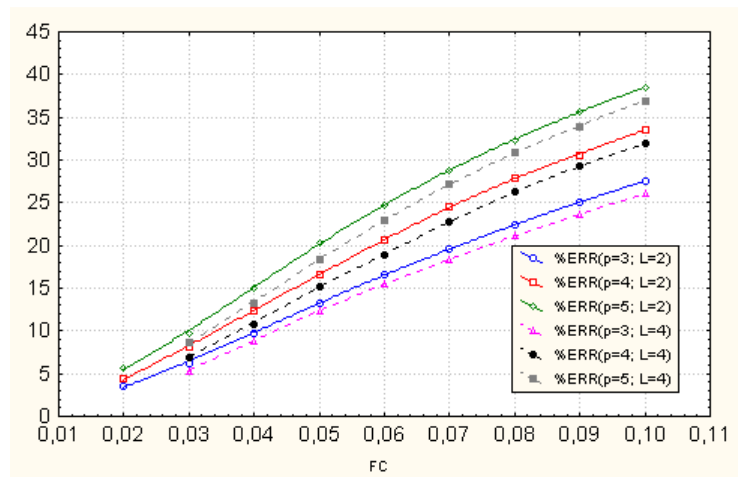


figure III.18 – taux d'erreur de classification (en %)

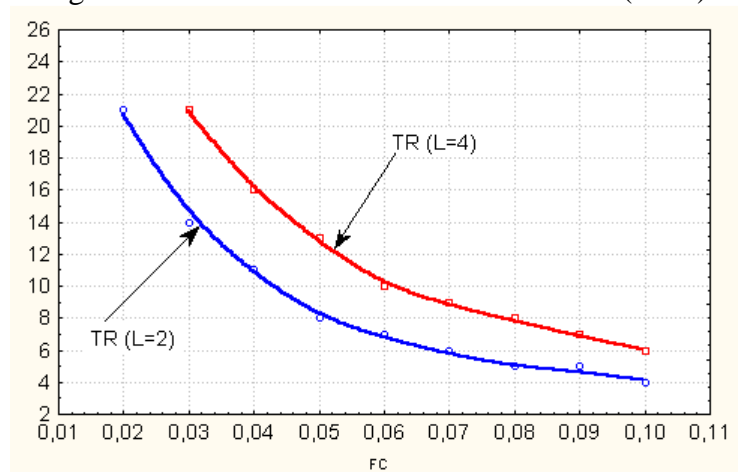


figure III.19 - temps de réponse en fonction de FC et L

Evidemment, le taux d'erreur de classification dépend aussi de l'amplitude du saut. Dans le tableau III.4 on donne, pour la même configuration expérimentale que dans le cas précédent, les valeurs du taux d'erreur de classification en fonction de l'amplitude du saut pour des filtres passe-bas d'ordre L=2 uniquement.

tableau III.4– le taux d'erreur de classification (en %)

Amp	FC	p = 3	p = 4	p = 5	TR
1.5	-	44.48	51.03	56.87	0
	0.02	1.45	1.66	1.87	21
	0.03	1.77	1.82	1.93	14
	0.04	2.36	2.55	2.70	11
	0.05	3.91	4.02	4.15	8
	0.06	5.42	5.54	6.10	7
	0.07	7.17	7.33	8.62	6
	0.08	8.81	9.15	11.32	5
	0.09	10.34	11.41	13.66	5
	0.1	12.02	13.16	16.03	4
2	-	32.38	37.96	43.05	0
	0.02	1.37	1.42	1.53	21
	0.03	1.07	1.16	1.21	14
	0.04	1.01	1.12	1.15	11
	0.05	1.07	1.06	1.12	8
	0.06	1.24	1.30	1.43	7
	0.07	1.65	1.76	2.10	6
	0.08	2.14	2.18	2.96	5
	0.09	2.83	2.98	3.81	5
	0.1	3.80	4.21	5.37	4
2.5	-	22.97	25.28	30.87	0
	0.02	1.10	1.15	1.31	21
	0.03	0.97	1.06	1.12	14
	0.04	0.86	0.94	0.97	11
	0.05	0.83	0.87	0.91	8
	0.06	0.81	0.85	0.87	7
	0.07	0.80	0.81	1.00	6
	0.08	0.96	1.00	1.18	5
	0.09	1.12	1.14	1.38	5
	0.1	1.55	1.71	2.08	4
3	-	14.22	17.96	19.39	0
	0.02	0.88	0.93	1.14	21
	0.03	0.71	0.79	0.94	14
	0.04	0.63	0.67	0.71	11
	0.05	0.55	0.60	0.65	8
	0.06	0.52	0.62	0.63	7
	0.07	0.60	0.66	0.68	6
	0.08	0.64	0.72	0.79	5
	0.09	0.73	0.83	0.91	5
	0.1	0.84	0.92	1.07	4

On remarque, d'abord, que l'augmentation de l'ordre du filtre ne réduit pas beaucoup l'erreur de classification (tableau III.3 et figure III.18) mais qu'elle augmente significativement le temps de réponse (tableau III.3 et figure III.19). Donc, on peut conclure qu'il est préférable de choisir un filtre d'ordre $L=2$ car il présente un temps de réponse plus court par rapport aux autres filtres d'ordre supérieur. Dans la figure III.20 on peut remarquer l'influence de l'ordre du filtre sur la "qualité" du filtrage d'un signal constitué de 200 observations. On observe que pour la même fréquence de coupure (FC) l'augmentation de l'ordre du filtre a surtout comme effet le lissage de la caractéristique filtrée. La sortie du filtre présente moins d'oscillations dans ce cas.

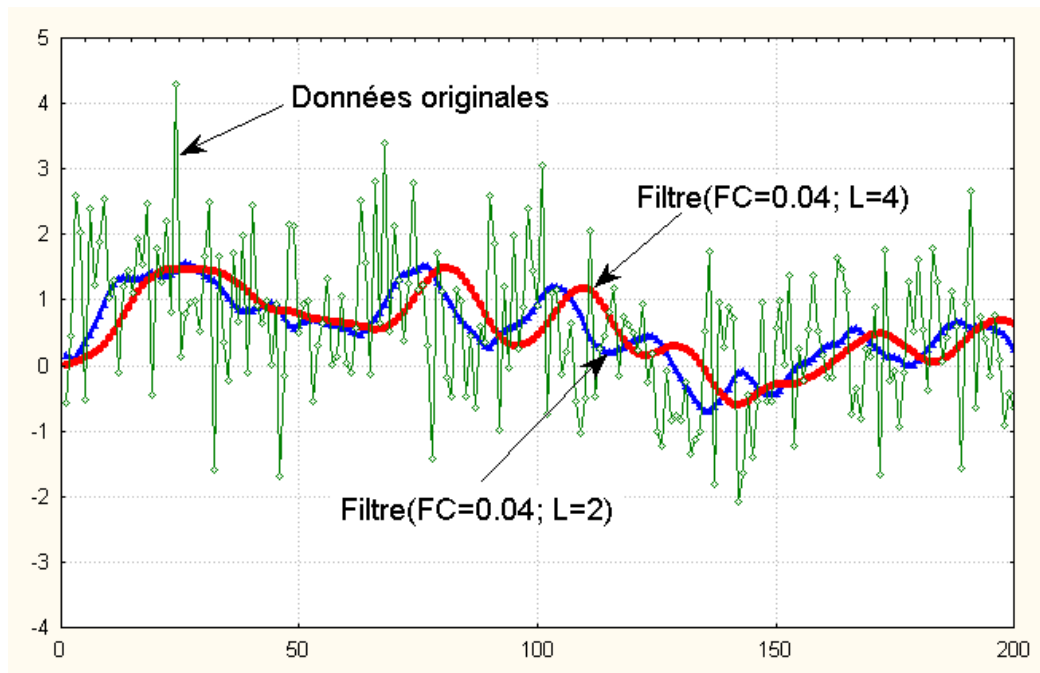


figure III.20 – Influence de l'ordre du filtre sur le filtrage

On peut conclure, en regardant les figures III.18 et III.19, que pour un saut d'amplitude 1σ :

- ✓ le temps de réponse des filtres augmente quand la fréquence de coupure diminue;
- ✓ pour la même fréquence de coupure, le temps de réponse augmente quand l'ordre du filtre augmente;
- ✓ l'erreur de classification augmente quand le nombre de variables augmente;
- ✓ l'erreur de classification diminue quand la fréquence de coupure diminue;
- ✓ pour la même fréquence de coupure, l'erreur de classification diminue quand l'ordre du filtre augmente;

Les mêmes conclusions sont également valables pour un saut d'amplitude 1.5σ (figure III.21.a). On peut conclure que, dans le cas des sauts de faible amplitude il est difficile de réaliser un compromis entre rapidité de détection et précision du diagnostic car celles-ci sont deux impératifs contradictoires. Nous pouvons remarquer que le problème soulevé ici est semblable à celui des limites de contrôle pour une carte de contrôle. Dans le cas des cartes de contrôle, si on veut détecter le dérèglement plus vite on

peut réduire la valeur de la limite de contrôle mais de cette manière on augmente également le risque d'avoir de fausses alertes. De même, si on veut détecter au plus vite, après le filtrage, des dérèglages de faible amplitude (inférieure à 1.5σ), on doit utiliser des filtre ayant une fréquence de coupure élevée mais ainsi on doit accepter le risque d'avoir un taux d'erreur important. Nous pensons qu'il est plus raisonnable, dans les applications pratiques, de commencer par la spécification d'un taux d'erreur maximum acceptable par l'utilisateur et ensuite de trouver le filtre qui permet d'atteindre cette valeur souhaitée. Ainsi, par exemple, si on veut détecter un dérèglement d'amplitude 1.5σ ayant un taux d'erreur inférieur à 2% , en regardant le tableau III.4 on voit qu'il faut utiliser un filtre de fréquence de coupure $FC=0.03$ qui présente un temps de réponse de 14 observations. Si ce temps de réponse est jugé trop long, on peut choisir un filtre ayant un temps de réponse de 8 observations ($FC=0.05$) mais dans ce cas, on aura un taux d'erreur de l'ordre de 4%! Il ne faut pas oublier, en même temps, que les valeurs présentées dans les deux tableaux III.3 et III.4 sont relatives à une situation concrète particulière et qu'elle ne sont données ici qu'à titre d'exemples. Néanmoins, les conclusions tirées après l'analyse de ces valeurs restent valables et généralisables pour toute situation pratique.

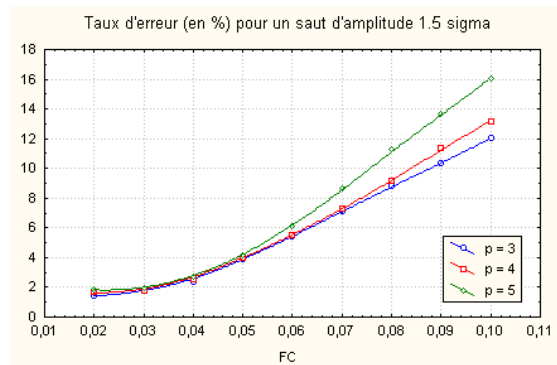


figure III.21.a

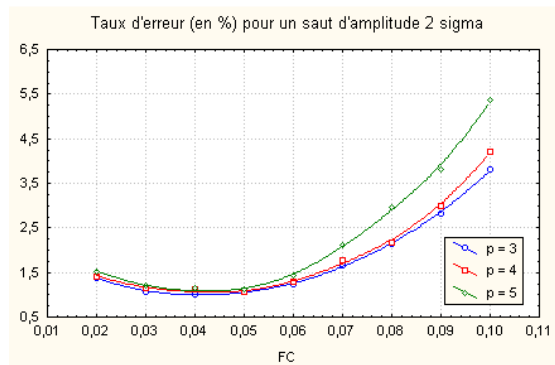


figure III.21.b

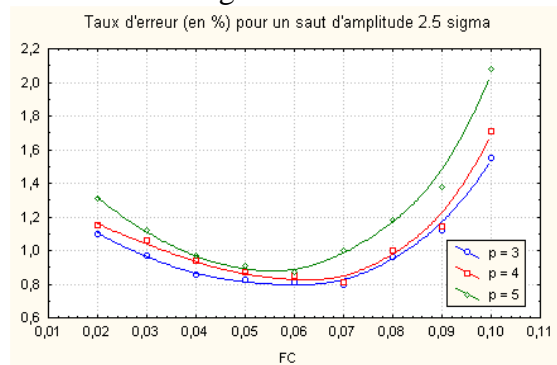


figure III.21.c

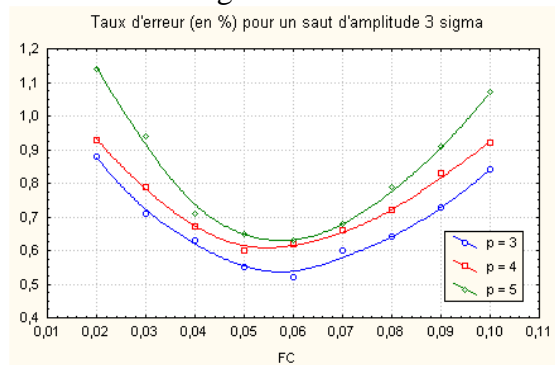


figure III.21.d

On remarque, également, en regardant les figures III.21.b, c et d que pour des dérèglages ayant une amplitude supérieure ou égale à 2σ la courbe du taux d'erreur change d'allure. On constate que la courbe ne présente plus une décroissance monotone mais qu'elle a une forme en "U" d'autant plus prononcée que l'amplitude du dérèglement est grande. Ce phénomène est dû à deux causes distinctes:

- ✓ le taux d'erreur est élevé à des fréquences de coupures réduites parce-que le filtre présente une grande "inertie" (figure III.15), ce qui veut dire une vitesse de

réaction faible face aux changements dans le signal. Toutes les erreurs de classification sont réalisées dans ces zones de transition dans le signal filtré.

- ✓ le taux d'erreur pour les fréquences de coupure élevées est important car la variation dans le signal reste grande et par conséquent il y a des observations appartenant à une certaine classe qui seront par erreur classées dans une autre classe.

On peut conclure, donc, que pour des dérèglages d'amplitude supérieure ou égale à 2σ il existe un filtre optimal, c'est-à-dire qui minimise le taux d'erreur de classement. La fréquence de coupure du filtre optimal est obtenue en déterminant le minimum sur la courbe du taux d'erreur en fonction de la fréquence de coupure.

3.4.4.4 Module de sélection des variables

Dans ce module on effectue le choix du nombre de variables susceptibles de mettre en évidence les différents modes de fonctionnement du processus. Ce choix repose sur une très bonne connaissance du processus physique étudié. Malheureusement, très souvent en pratique, on ne dispose pas d'une telle connaissance et on doit prendre en compte un très grand nombre de variables pour le suivi du processus. La conséquence principale d'une dimension trop élevée de l'espace des variables est qu'il est très difficile de trouver une bonne règle de décision pour effectuer le diagnostic du processus (Jain et Ramaswami [59]) car certaines variables peuvent être peu informatives, voire perturbatrices et si on les prend en compte on peut influencer d'une façon négative la qualité de la discrimination.

Mais, il faut éviter de trop réduire le nombre de variables car il y a le risque d'avoir de nouveaux modes de fonctionnement non-identifiables dans l'espace de variables réduit alors qu'ils auraient pu être détectés dans l'espace originel. Jousselin [61], Foroutan et Sklansky [31] proposent l'utilisation d'une règle d'arrêt basée sur la perte d'information admissible lors du processus de réduction.

Le pouvoir discriminant, par rapport aux classes, des sous-ensembles de variables est évalué à l'aide des critères prenant en compte la probabilité d'erreur de classification ou par des critères de séparation des classes (distances probabilistes et géométriques). D'ailleurs les deux critères sont équivalents car la probabilité d'erreur de classification est une mesure idéale de la séparation des classes. Pourtant, comme nous l'avons déjà précisé, la probabilité d'erreur est difficilement évaluable analytiquement et par conséquent, il faut l'estimer. S'il n'y a pas de bruit et si on dispose d'une quantité infinie d'observations, la probabilité d'erreur décroît quand le nombre des variables augmente. Comme, en pratique, on dispose d'échantillons de taille finie elle n'est pas monotone décroissante. Au début, les performances augmentent lorsque les nouvelles variables sont ajoutées mais à partir d'un certain moment l'ajout de variables supplémentaires dégrade les performances. Il existe donc un nombre de variables optimal. Un critère plus simple est d'analyser la séparation des classes à l'aide de la matrice de variance inter-classe (**B**) et intra-classe (**W**) (voir le paragraphe 3.2.1 pour une définition plus détaillée de ces deux matrices). Deux indicateurs ont été proposés dans ce but. Ils sont de la forme suivante:

$$J_1 = \text{tr} \left(\frac{S_2}{S_1} \right) \text{ et } J_2 = \log \frac{|S_2|}{|S_1|} \quad (3.22)$$

avec (S_1, S_2) des matrices qui peuvent être remplacées par les combinaisons suivantes: (\mathbf{T}, \mathbf{B}) , (\mathbf{W}, \mathbf{B}) et (\mathbf{W}, \mathbf{T}) , où \mathbf{T} est la matrice de variance-covariance totale. L'avantage de ces indicateurs est qu'il sont en rapport direct avec la géométrie des classes et qu'ils peuvent être calculés à l'aide de formules récursives, d'où un gain de temps. Cependant, ils ne sont pas directement liés à la probabilité d'erreur de classification et ils ne peuvent pas donner des indications sur le recouvrement des classes. Un autre inconvénient de ces critères est que dans le cas où les distributions sont significativement non normales ils ne permettent pas de sélectionner les variables nécessaires à la préservation de structures complexes (Fukunaga et Mantock [35]; Foroutan et Slansky [31], Fukunaga [34]).

Il est donc évident que le choix du nombre des variables reste un problème délicat, d'autant plus qu'il n'existe pas de procédure automatique permettant de le résoudre. En pratique, le plus souvent, on procède par essai/erreur, c'est-à-dire qu'un certain nombre de variables est choisi arbitrairement et la validité du modèle ainsi obtenu est testée. Le seul critère de validité est la possibilité d'interprétation des classes obtenues comme des modes de fonctionnement du processus. Cela reste d'autant plus difficile que la quantité d'informations sur la structure en classes que l'on cherche à mettre en évidence est réduite. En pratique, il est préférable de prendre en compte un nombre réduit de variables quand le nombre des données est très faible.

Une autre méthode est celle proposée par McCabe [84] qui consiste à considérer toutes les partitions possibles de l'ensemble de \mathbf{p} variables en deux classes: la classe des variables retenues (de taille \mathbf{p}_1) et la classe des variables rejetées (de taille $\mathbf{p}_2 = \mathbf{p} - \mathbf{p}_1$). Par la suite, la matrice de variance-covariance de l'ensemble complet de variables (\mathbf{S}) est à son tour divisée en quatre composantes:

$$\mathbf{S} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \quad (3.23)$$

où: S_{11} est la matrice de variance-covariance pour les \mathbf{p}_1 variables retenues et S_{22} est la matrice de variance-covariance pour les \mathbf{p}_2 variables rejetées.

Deux critères sont intéressants pour l'identification du nombre des variables retenues:

- 1) $\max |S_{11}|$, où "||" est l'opérateur déterminant
- 2) $\min [\text{tr}(S_{22} - S_{21}S_{11}^{-1}S_{21}^T)]$, où tr est l'opérateur "trace"

Le premier critère permet de retenir le maximum de variation dans les données lorsque la matrice de variance-covariance est retenue. Dans le cas où la matrice de corrélation est retenue, le critère permet de retenir uniquement les caractères les moins corrélés. De plus, ce critère peut être facilement implémenté en utilisant des formules récurrentes éliminant ainsi le calcul du déterminant des matrices de grandes dimensions à chaque fois.

Le deuxième critère correspond à l'information perdue en éliminant le sous-ensemble de variables \mathbf{p}_2 . Pour ce critère il n'existe pas des formules récurrentes simples, ni d'algorithmes efficaces pour examiner un grand nombre de solutions.

Si, par une inspection visuelle des données, on observe une structure en classe alors on peut penser que le choix des variables est bon. Représenter les données multidimensionnelles en conservant toutes les dimensions pour éviter une perte d'information reste un objectif assez difficile. Plusieurs techniques ont été développées pour obtenir une représentation graphique des données initiales dans un espace de dimensions réduites. Ces méthodes portent le nom de "**méthodes de projection**". La représentation des données issues des méthodes de projection est particulièrement utile pour localiser des classes présentes dans un ensemble de données. L'inspection de ces projections permet de conclure quant à la validité du choix des variables. Cependant, ces méthodes présentent un certain nombre d'inconvénients parmi lesquels nous mentionnons: la forme de la disposition réelle des données dans l'espace multidimensionnel est déformées par la projection; elles reposent sur l'hypothèse que la dimension intrinsèque des données permet la représentation dans des plans, ce qui n'est pas toujours vrai; elles nécessitent l'intervention de l'analyste pour l'inspection visuelle des projections.

Malgré tout, il est très peu probable de ne sélectionner que des variables qui ne mettent pas en évidence des modes de fonctionnement. Par conséquent, le problème est plutôt un problème de sélection de variables, c'est-à-dire d'élimination des variables redondantes ou non informatives.

3.4.4.5 Module de classification

3.4.4.5.1 Introduction

L'objectif d'une classification est d'obtenir des classes constituées d'éléments qui se ressemblent beaucoup plus entre eux que des éléments pris dans des classes différentes. A partir des années 1960, suite à l'évolution enregistrée dans le monde informatique, un grand nombre d'algorithmes automatisant les classifications a été créé, d'où le terme usuel de "classification automatique", transposition du terme anglais "cluster analysis".

La réalisation des classes peut être **un but** (obtenir une répartition des individus en groupes homogènes et bien séparés), **un moyen** (réduire le volume de données car beaucoup de méthodes sont inadaptées au traitement de grandes quantités de données à cause du coût de calcul prohibitif ou du stockage en mémoire dépassant la capacité de l'ordinateur) **une suite** (c'est le cas de la méthode FNAD où elle est utilisée comme une analyse complémentaire avant l'analyse discriminante; elle peut également être utilisée après une analyse en composantes principales ou une analyse factorielle de correspondances pour aider à l'interprétation des graphiques de projection). Nous avons utilisé les techniques de classification comme un moyen pour la réalisation de la bibliothèque des classes, bibliothèque qui servira, par la suite, au module d'analyse discriminante située dans la chaîne directe (figure III.4).

Une "bonne" classification doit satisfaire à deux conditions: **cohésion** (ou homogénéité à l'intérieur des classes) et **isolement** (ou hétérogénéité entre les classes). Ceci signifie que les individus d'une même classe doivent se ressembler au maximum tandis que ceux de deux classes différentes doivent être les plus différents possible. De ce point de vue, l'objectif des techniques de classification ressemble à l'objectif descriptif de l'analyse discriminante, à part le fait que dans le cas de l'analyse

discriminante les classes sont connues a priori. Tout comme dans le cas de l'analyse discriminante, on estime la dissemblance entre deux individus par une mesure de dissimilitude ou de distance. Les individus qui se ressemblent le plus sont ceux qui ont la plus faible distance ou dissimilitude, calculée pour toutes les variables. Une dissimilitude doit satisfaire pour tout individu x_1, x_2, x_3 deux conditions:

$$\begin{aligned} d(x_1, x_2) &= d(x_2, x_1) \geq 0 \\ d(x_1, x_2) &= 0 \Leftrightarrow x_1 = x_2 \end{aligned} \quad (3.24)$$

Si elle vérifie en plus: $d(x_1, x_3) \leq d(x_1, x_2) + d(x_2, x_3)$ alors cette mesure est une distance. Il existe de nombreuses formules de calcul de distances; nous donnons ici celles qui sont les plus utilisées:

i. *la distance euclidienne au carré*: $d_{ik}^2 = \sum_{j=1}^p (x_{ij} - x_{kj})^2$, où x_{ij} est la valeur de l'individu x_i mesurée pour la variable j , et p est le nombre des variables;

ii. *la distance euclidienne*: $d_{ik} = \sqrt{\sum_{j=1}^p (x_{ij} - x_{kj})^2}$

iii. *la distance euclidienne pondérée*: $d_{ik}^2 = \sum_{j=1}^p m_j (x_{ij} - x_{kj})^2$, où m_j est le poids associé à la variable j

iv. *la distance du khi2 (χ^2)*: $d_{ik}^2 = \frac{\sum_{j=1}^p \left[\frac{x_{ij}}{x_{i\bullet}} - \frac{x_{kj}}{x_{k\bullet}} \right]^2}{x_{\bullet j}}$, où $x_{i\bullet} = \sum_{j=1}^p x_{ij}$,

$x_{k\bullet} = \sum_{j=1}^p x_{kj}$, et $x_{\bullet j} = \sum_{i=1}^n x_{ij}$, n étant le nombre d'individus ou d'observations.

v. *la distance rectangulaire*: $d_{ik} = \sum_{j=1}^p m_j |x_{ij} - x_{kj}|$

Le résultat d'une classification est: **une partition** (la liste d'individus qui constituent chaque classe; c'est la représentation la plus simple) ou **une hiérarchie** (représentation plus complexe des partitions emboîtés et qui se présente graphiquement sous la forme d'un arbre hiérarchique ou *dendrogramme*). Certains algorithmes fournissent directement une partition et portent le nom d'**algorithmes non-hiérarchiques**, d'autres fournissent successivement une hiérarchie puis une partition et portent le nom d'**algorithmes hiérarchiques**.

Algorithmes de classification non-hiérarchiques

Les classifications non hiérarchiques sont basées sur le principe d'une partition initiale qui est "améliorée" au cours d'itérations successives. Nous décrirons d'abord la méthode des centres mobiles et ensuite la méthode des "nuées-dynamiques" qui est une variante de la précédente.

L'algorithme d'agrégation autour des centres mobiles

Le déroulement de cet algorithme est le suivant: dans un premier temps on regroupe les individus autour de k centres arbitraires $g_1^0, g_2^0, \dots, g_k^0$ de la manière suivante: la classe associée à g_j^0 est constituée de l'ensemble d'individus plus proches de g_j^0 que de tout autre centre. Géométriquement ceci revient à partager l'espace des individus en k zones définies par les plans médiateurs des segments $g_i^0 g_j^0$. Le choix du nombre de centres k est délicat car il n'y a aucune règle pour fixer ce nombre. Nous verrons plus loin que l'examen des "formes fortes" peut permettre de résoudre ce problème. On calcule ensuite pour chaque classe formée le centre de gravité (où individu moyen). On obtient ainsi un nombre de k nouveaux centres de gravité $g_1^1, g_2^1, \dots, g_k^1$. On effectue alors une deuxième partition en regroupant les individus autour des centres g_j^1 qui prennent alors la place des centres g_j^0 de la première étape. On calcule les centres de gravité $g_1^2, g_2^2, \dots, g_k^2$ de ces nouvelles classes, on regroupe les individus autour d'eux et ainsi de suite jusqu'à ce que la qualité de la partition mesurée par l'inertie intra-classe (ou la variance intra-classe) ne s'améliore plus. On peut démontrer que le fait de passer d'un centre arbitraire au centre de gravité diminue l'inertie intra-classe et fournit donc des classes plus homogènes. Si le nombre d'individus à classer n'est pas trop élevé, l'expérience montre que cette "convergence" est très rapide: une dizaine d'itérations sont en général suffisantes. Pour une définition plus détaillée de la variance intra-classe et de la variance inter-classe, le lecteur est invité à lire les notions présentées au paragraphe 3.2 dédié à l'analyse discriminante. Atteindre la partition finale ne signifie pas que l'on ait atteint la meilleure partition possible au sens du critère de l'inertie intra-classes minimum. En effet, une meilleure partition finale pourra être obtenue avec une partition initiale différente. Le résultat final obtenu est dépendant du choix de la partition initiale et ceci est l'un des inconvénients majeurs de la méthode. Un autre inconvénient de la méthode est lié au risque d'aboutir sur à un nombre de classes inférieur à k . Il faut tenir compte aussi du fait que la comparaison des variances intra-classes de deux partitions différentes ne peut se faire que pour des partitions ayant le même nombre de classes car, plus le nombre de classes est élevé et plus la variance intra-classes est faible (sinon, la partition idéale serait celle où toutes les classes ne contiendraient qu'un seul individu).

Sans aucune information sur les données, on effectue plusieurs fois l'algorithme avec un nombre différent de classes et on compare les résultats pour comprendre le mécanisme de regroupement des individus. On peut également, pour le même nombre de classes, exécuter plusieurs fois l'algorithme avec des partitions initiales différentes et comparer ensuite les différentes partitions finales. Entre deux partitions possédant le même nombre de classes, on pourra retenir celle minimisant la variance intra-classe. Lorsque deux ou plusieurs individus ont toujours été réunis dans la même classe au cours des différentes exécutions pour des partitions initiales différentes, ces individus forment un groupe particulièrement robuste: on l'appelle **forme forte**. La partition que l'on pourra retenir et examiner est la réunion de toutes ces formes fortes qui constituent des classes très homogènes. Ceci permet également de ne plus se préoccuper du choix a priori du nombre des classes, le nombre des formes fortes ne dépendant pas directement du nombre de classes a priori.

La méthode des nuées dynamiques

Cette méthode est une généralisation de la méthode d'agrégation autour des centres mobiles. La différence fondamentale est la suivante: au lieu de définir une classe par un seul point, son centre, qui ne peut pas être un des individus de l'ensemble à classer, on la définit par q individus formant un "**noyau**" qui, s'ils sont bien choisis, seront plus représentatifs de la classe qu'un simple centre de gravité. Ces noyaux permettront par la suite d'interpréter les classes. L'algorithme génère, dans une première étape, une partition en regroupant les individus autour des noyaux arbitrairement choisis (un individu est alloué au noyau le plus proche) et ensuite, calcule les nouveaux noyaux représentatifs des classes ainsi formées et ainsi de suite jusqu'à ce que la qualité de la partition ne s'améliore plus. Trois fonctions sont nécessaires pour implémenter cet algorithme:

- ✓ la première qui calcule la distance entre un individu et un noyau;
- ✓ la deuxième qui à une partition en k classes associe les k noyaux de q points, représentatifs de ces classes;
- ✓ la troisième qui mesure la qualité d'une partition;

Connaissant ces trois fonctions, le nombre de classes et l'effectif des noyaux, l'algorithme est entièrement déterminé.

Algorithmes de classification hiérarchique

Il existe deux familles d'algorithmes de classification hiérarchique: les *algorithmes ascendants* (ou encore agglomératifs) qui procèdent à la construction de classes par agglomérations successives d'objets deux à deux et qui fournissent une hiérarchie de partitions des objets; les *algorithmes descendants* (ou encore divisifs) qui procèdent par dichotomies successives de l'ensemble d'objets fournissant ainsi une hiérarchie de partitions.

L'algorithme de classification hiérarchique ascendante

Le principe de l'algorithme est le suivant: dans une étape préliminaire, on calcule les distances de tous les individus deux à deux, ensuite, dans la première étape, on cherche la distance la plus faible et on forme une première partition en agrégeant les individus les plus proches; dans la deuxième étape on calcule les distances entre le groupe qui vient d'être formé et tout le reste des individus (la formule de calcul de la distance est appelée **critère d'agrégation**); dans la troisième étape on recherche de nouveau la distance la plus faible et on procède à une nouvelle agrégation et ainsi de suite jusqu'à ce que tous les individus aient été regroupés dans une classe unique.

Les critères d'agrégation sont les différentes formules de calcul de la distance entre un groupe quelconque G obtenu par fusion de deux sous-groupes G_1 et G_2 et un autre groupe Q (les groupes peuvent être réduits à un seul élément). Parmi les critères d'agrégation les plus utilisés on peut citer:

- ✓ le critère d'agrégation de la "**distance minimale**" ou du "lien simple" (en anglais "single linkage"): la distance est la plus petite des distances de Q avec G_1 ou avec G_2 :

$$d(Q, G) = \min \{d(Q, G_1), d(Q, G_2)\}$$

L'avantage de ce critère est qu'il sépare bien les groupes et qu'il est capable de détecter des classes de forme irrégulière ou allongée. Son inconvénient est le "chaînage" (par agglomérations successives des points très différents arrivent à fusionner et les niveaux de regroupement se retrouvent régulièrement échelonnés, ce qui ne permet pas d'identifier des groupes parmi les individus).

- ✓ le critère d'agrégation de la "**distance maximale**" ou du "lien complet" (en anglais "complete linkage"): la distance est la plus grande des distances de Q avec G_1 ou avec G_2 :

$$d(Q, G) = \max \{d(Q, G_1), d(Q, G_2)\}$$

Ce critère convient très bien dans le cas des classes compactes de tailles à peu près égales. Les inconvénients de ce critère sont la sensibilité aux valeurs extrêmes ou aberrantes et la tendance à fusionner les classes à faibles effectifs.

- ✓ le critère d'agrégation de la "**moyenne des distances pondérées**" ou du "lien moyen" (en anglais "average linkage"): la distance est pondérée par le nombre d'individus dans chacun des sous-groupes:

$$d(Q, G) = [m_1d(Q, G_1) + m_2d(Q, G_2)]/(m_1 + m_2)$$

où m_1 représente le nombre d'individus dans le groupe G_1 , m_2 représente le nombre d'individus dans le groupe G_2 ;

Ce critère convient bien dans le cas des classes compactes et de tailles à peu près égales. Son inconvénient est qu'il a tendance à fusionner les classes à faible variance et à produire des classes de variances égales.

- ✓ le critère d'agrégation du "**moment d'ordre deux**":

$$d(Q, G) = [(m_1+m_2)d(Q,G_1) + (m_1+m_2)d(Q,G_2) - md(G_1,G_2)]/(m+m_1+m_2)$$

où m est le nombre d'individus du groupe Q.

Comme le critère de la moyenne, ce critère est très bien adapté aux classes compactes de tailles à peu près égales. Il a tendance à fusionner les classes à faibles effectifs. Il est sensible aux valeurs extrêmes ou aberrantes.

- ✓ le critère de l'inertie ou "**la méthode de Ward**": cherche à optimiser, à chaque étape, selon des critères liés à des calculs d'inertie (ou de la variance), la partition obtenue par agrégation de deux éléments. Une "bonne" partition est celle pour laquelle l'inertie inter-classe est forte et donc, par conséquent, l'inertie intra-classe faible. Comme lorsque l'on passe d'une partition en $k + 1$ classes à une partition en k classes en regroupant deux classes en une seule, l'inertie inter-classe ne peut que diminuer, le critère de regroupement sera: fusionner les deux classes pour lesquelles la perte d'inertie est la plus faible. Ceci revient à

réunir les deux classes les plus proches en prenant comme distance entre deux classes la perte d'inertie qui résulte du regroupement.

L'algorithme de classification hiérarchique ascendante ne fournit pas une partition en k classes d'un ensemble de n individus (comme les algorithmes non-hiérarchiques) mais une hiérarchie de partitions, se présentant sous la forme d'un arbre (appelé aussi dendrogramme) et contenant $n-1$ partitions. L'intérêt de cet arbre est qu'il peut donner une idée du nombre de classes existant effectivement dans la population. L'arbre est "coupé" à un niveau donné et on identifie les groupes (ou classes) constitués avant cette coupure. L'ensemble des classes formées à ce niveau constitue une partition.

L'algorithme de classification hiérarchique descendante

Le principe de la classification hiérarchique descendante est de diviser successivement chaque classe en deux sous-classes, de tailles non obligatoirement identiques. L'algorithme part d'une classe unique contenant tous les individus pour aboutir après plusieurs étapes à un individu par classe. Il détermine à chaque étape la classe à diviser et deux "points-noyaux" parmi les individus, à partir desquels sera réalisée la division en se basant sur la dispersion à l'intérieur des classes. Les algorithmes hiérarchiques descendants sont assez peu utilisés maintenant car ils fournissent la même représentation des données que les algorithmes hiérarchiques ascendants (un arbre hiérarchique), mais ne possèdent pas plus d'intérêt. Par contre, ce type de méthode nécessite beaucoup de temps de calcul car il faut, à chaque étape, envisager toutes les combinaisons possibles.

Les deux algorithmes non-hiérarchiques présentés (la méthode d'agrégation autour des centres mobiles et la méthode des nuées dynamiques) permettent de traiter rapidement de grands ensembles d'individus mais ils supposent que le nombre des classes k est fixé. Si ce nombre ne correspond pas à la configuration véritable du nuage des individus on risque d'obtenir des partitions de valeur douteuse. Pour remédier à ce problème on peut essayer diverses valeurs de k , mais dans ce cas le temps de calcul sera augmenté aussi considérablement. Lorsque le nombre d'individus n'est pas trop élevé on aura plutôt recours à des méthodes hiérarchiques qui ne nécessitent pas la spécification a priori du nombre de classes k . Les partitions obtenues à partir d'algorithmes différents ne sont pas forcément identiques et il n'y a aucun critère mathématique permettant de décider si une partition est meilleure qu'une autre. On peut, bien sûr, analyser la bonne séparation des classes ou leur bonne cohésion mais ces critères ne suffisent pas pour savoir si une partition est meilleure qu'une autre. C'est pour cette raison que, dans le module d'analyse "off-line" de la méthode FNAD, nous proposons une combinaison de deux algorithmes de classification: les algorithmes non-hiérarchiques et les algorithmes hiérarchiques. Cette combinaison nous semble beaucoup plus appropriée pour ce type d'application où le volume des données peut être important et, en même temps, l'information concernant le nombre réel de classes inexistante. Il s'agit donc d'une méthode de **classification mixte** qui va être décrite dans le paragraphe suivant.

3.4.4.5.2 Méthode de classification mixte

L'algorithme de *classification mixte* s'effectue en trois étapes:

- ✓ étape 1: l'ensemble des éléments à classer subit un partitionnement initial (autour des centres mobiles ou nuées dynamiques) de façon à obtenir quelques dizaines, voir quelques centaines de groupes homogènes;
- ✓ étape 2: on applique une agrégation hiérarchique de ces groupes formés à l'étape 1, dont le dendrogramme suggérera éventuellement le nombre de classes finales à retenir;
- ✓ étape 3: on effectue une optimisation finale (par la technique des centres mobiles ou nuées dynamiques) pour les partitions obtenues suite à la coupure de l'arbre à l'étape 2;

La figure III.22 schématise les différentes étapes de l'algorithme de classification mixte.

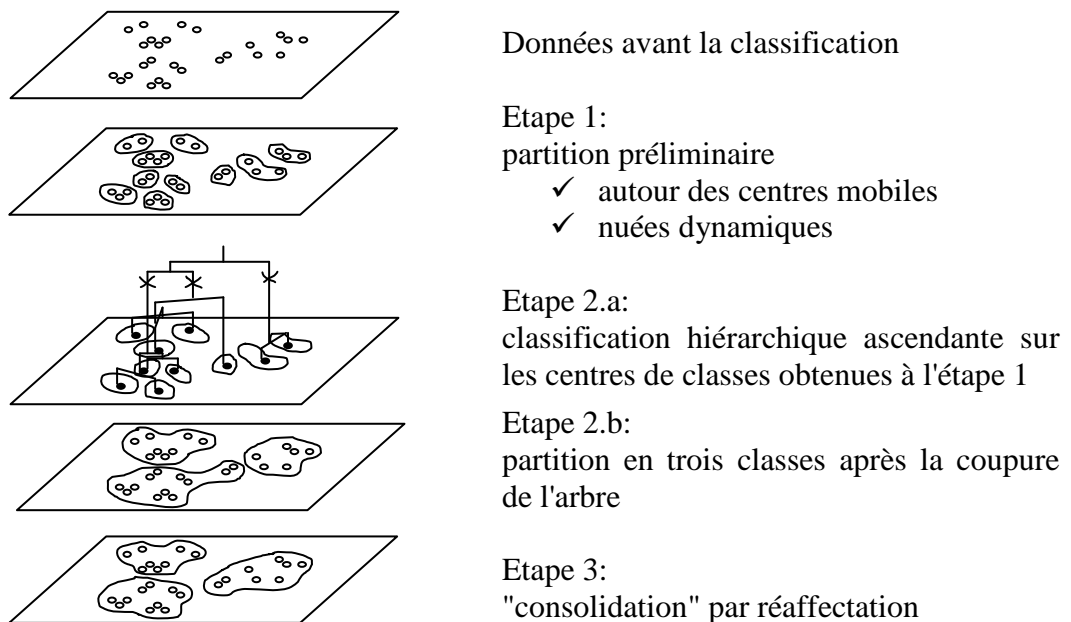


figure III.22 - Schéma de la méthode de classification mixte

La première étape aide à l'obtention rapide et à un faible coût, de k classes homogènes, où k est largement plus élevé que le nombre réel des classes dans la structure des données. L'algorithme d'agrégation autour des centres mobiles ou l'algorithme des nuées dynamiques peuvent être utilisés à cette étape préliminaire. Ainsi, l'inertie entre les classes est augmentée après chaque itération de l'algorithme et au final on obtient un nombre fixé au préalable de classes mais qui dépend du choix initial des centres. On ne peut pas atteindre de cette manière la partition optimale mais en utilisant les formes fortes on peut néanmoins améliorer la partition obtenue.

La seconde étape consiste à effectuer une classification hiérarchique ascendante où les éléments terminaux de l'arbre sont les k centres de classe de la partition obtenue à l'étape précédente. Le but de l'étape d'agrégation hiérarchique est de reconstituer les classes qui ont été fragmentées et d'agréger des éléments apparemment dispersés autour de leurs centres d'origine. Il est très possible que des groupes soient proches les uns des autres. Ils correspondent, en effet, à un groupe "réel" qui a été artificiellement divisé à

l'étape 1. Le choix du niveau de la coupure de l'arbre peut être facilité par une inspection visuelle. En coupant l'arbre au niveau d'un saut important de l'indice de niveau (indice qui donne la distance entre deux classes ou groupes) on peut obtenir une partition de "bonne" qualité. Evidemment, en pratique, la situation n'est pas toujours facile, car souvent on ne peut pas distinguer clairement par une simple inspection visuelle une zone de coupure nette. Dans ce cas l'utilisateur peut choisir entre deux ou trois niveaux de coupure possibles et donc entre deux ou trois partitions finales. La coupure de l'arbre peut être facilitée par l'examen de l'histogramme des indices croissants de niveau et l'on coupera au niveau pour lequel cet histogramme marque un palier important.

La partition finale est obtenue, après la coupure de l'arbre, par une nouvelle réaffectation des individus. Cette nouvelle réaffectation autour des centres de gravités prédéterminés à l'étape antérieure a pour but d'homogénéiser encore plus les classes. Après un certain nombre d'itérations, il n'y a plus de réaffectation à opérer et le calcul est arrêté. En pratique, la procédure est arrêtée dès que l'inertie entre les classes cesse de croître de façon sensible.

3.4.5 Module bibliothèque de classes

La bibliothèque de classes résume les caractéristiques des différents modes de fonctionnement du processus (le fonctionnement sous-contrôle et le fonctionnement hors-contrôle). Selon le degré de connaissance des ces différents modes de fonctionnement, la bibliothèque peut être complète ou incomplète. Si tous les états possibles (de bon fonctionnement et de mauvais fonctionnement) ont été recensés, alors la bibliothèque est dite complète. Elle englobe dans ses différentes classes les paramètres (vecteur moyenne et matrice de variance-covariance, par exemple dans le cas gaussien) caractérisant tous les états identifiables du système. Dans ce cas, le schéma de la méthode FNAD (figure III.4) peut être simplifié, en éliminant le module de rejet situé dans la chaîne directe. Si l'expérimentation de tous les états n'est pas possible, alors la bibliothèque est dite incomplète. Dans ce cas, elle contient seulement les classes correspondant aux états naturellement et facilement observables, et par conséquent dans la chaîne directe on a besoin d'un module de rejet. La construction de la bibliothèque de classes ne se résume pas à une simple construction d'une partition. Elle nécessite également l'identification des attributs principaux des classes pour pouvoir les interpréter physiquement (Simon [109]). Si les classes n'ont aucune interprétation physique alors il faut renoncer à la solution obtenue car elle est inutilisable pour le diagnostic.

Au cours de la phase d'exploitation, les nouvelles observations collectées sur le système sont, soit ajoutées à un mode de fonctionnement connu, soit rejetées à l'aide du module de rejet situé dans la chaîne directe et déposées temporairement dans une mémoire tampon. Ces observations rejetées ou associées aux classes doivent être prises en compte pour l'actualisation de la bibliothèque de classes. Cette actualisation comporte deux étapes:

- ✓ mise à jour des paramètres des classes existantes: Nous avons vu que, dans le cas gaussien, les classes composant la bibliothèque de classes sont représentées par leurs paramètres caractéristiques comme le vecteur de moyennes ou la matrice de variance-covariance. Une manière simple de procéder est de modifier les paramètres de la classe chaque fois qu'un nouveau point est affecté à cette

classe. Quesenberry [99] propose les formules des mise à jour suivantes pour le vecteur de moyennes et pour la matrice de variance-covariance:

$$\bar{X}_j = \frac{1}{j} [(j-1)\bar{X}_{j-1} + X_j]$$

où: $\bar{X}_{j-1} = \frac{1}{j-1} \sum_{i=1}^{j-1} X_i$ est le vecteur des moyennes pour les observations X_i (distribuées selon une distribution p-normale de moyenne μ et matrice de variance-covariance Σ) comprises entre $i = 1$ et $i = j-1$, et peut être considéré comme un estimateur du vecteur réel de moyennes μ du processus. ($\bar{X}_0 = 0_p$)

$$S_j = \frac{1}{j-1} Q_j$$

où: S_j est l'estimateur de la matrice de variance-covariance pour les observations de 1 à j

$$Q_j = Q_{j-1} + \frac{j-1}{j} d_j d_j^T \quad (\text{avec } Q_0 = 0_{p \times p})$$

et $d_j = X_j - \bar{X}_{j-1}$ est l'écart de l'observation présente j par rapport au vecteur de moyennes estimé à partir de $j-1$ observations antérieures.

Cette actualisation permanente est simple à réaliser avec un algorithme de calcul qui ne nécessite pas une programmation très élaborée. Une deuxième possibilité est de faire une mise à jour différée (quand un certain nombre de points ont été affectés à la même classe). Smolarz [110] propose de régler cette valeur en fonction de la dynamique du processus par apprentissage.

- ✓ création de nouvelles classes: Elle doit intervenir lorsque les point rejetés par rejet de distance forment une ou plusieurs nouvelle(s) classe(s). L'apparition d'une nouvelle classe peut se faire d'une manière soudaine (un changement d'état ou une défaillance du processus physique) ou suite à une transition lente (dérive de capteur, vieillissement d'un outil, outillage, etc...). La procédure retenue dans le cadre de la méthode FNAD pour la création de nouvelles classes est la suivante: lorsque un certain nombre fixé de points rejetés est atteint, on effectue leur analyse à l'aide du module de classification en commençant par une classification hiérarchique ascendante (car le nombre d'observations rejetées est supposé relativement réduit) suivie d'une éventuelle homogénéisation des classes à l'aide d'une classification non-hiérarchique (si plusieurs classes ont été détectées à l'étape précédente). Après la création d'une nouvelle classe, il se peut que de nouvelles variables soient nécessaires pour la discriminer par rapport aux autres classes. Jousselin [61] propose une procédure de recherche exhaustive pour augmenter la dimension de l'espace courant. Si on veut ajouter le minimum de variables, il n'y a pas un grand inconvénient à parcourir toutes les solutions (pour accélérer cette opération, il serait possible de considérer en premier les attributs de la nouvelle classe formée).

3.4.6 Module d'analyse "on-line"

Le module d'analyse "on-line" est utilisé pour effectuer le diagnostic en temps réel du fonctionnement du processus. Il se compose de trois sous-modules: le module de filtrage numérique, le module de rejet et le module d'analyse discriminante. Nous avons déjà présenté le module de filtrage numérique dans le paragraphe 3.4.4.3 et, par conséquent, nous n'allons détailler dans ce paragraphe que les modules de rejet et d'analyse discriminante.

3.4.6.1 Module de rejet

En pratique, il est très probable que toutes les classes de fonctionnement du processus ne soient pas identifiées dès le début d'où la nécessité de ne pas construire un système de décision basé uniquement sur cette connaissance imparfaite.

Le module de rejet permet de ne pas classer à tout prix une observation dans une des classes déjà connues et d'éviter ainsi des décisions erronées. Il y a deux situations où le rejet d'une observation est nécessaire:

- ✓ quand une observation ne peut pas être affectée avec une grande probabilité (significative du point de vue statistique) à une des classes existantes dans la bibliothèque de classes. Ce type de rejet introduit par Chow [13] porte le nom de "**rejet d'ambiguïté**".
- ✓ quand une observation ne peut pas être associée aux classes déjà connues car elle est due à l'apparition d'une nouvelle classe (Dubuisson [28]). Ce type de rejet porte le nom de "**rejet de distance**". Ainsi, on peut améliorer la connaissance existante en intégrant dans la bibliothèque des classes des nouvelles classes qui n'ont pas été détectées pendant l'analyse de l'existant.

3.4.6.2 Module d'analyse discriminante

Comme on peut le constater dans la figure III.4, le module d'analyse discriminante est situé dans la chaîne directe et est utilisé afin de réaliser un diagnostic en temps réel de l'état de fonctionnement du processus. Les données multivariées sont filtrées à l'aide du module de filtrage numérique avant d'être soumises à l'analyse discriminante qui, par la suite, permettra de déterminer quelle est la variable responsable du dérèglement dans le cas du fonctionnement anormal du processus.

Pour expliciter le fonctionnement du module d'analyse discriminante et la manière dont le diagnostic est réalisé, nous allons présenter le cas d'une analyse de trois variables. L'applicabilité de la méthode FNAD ne se limite pas, bien évidemment, à ce cas particulier. Nous avons choisi d'utiliser trois variables car elles définissent un espace à trois dimensions qui est plus facile à imaginer (moins abstrait qu'un espace de dimension plus élevée) et, en plus, on a l'avantage de pouvoir le représenter sous une forme graphique. Nous avons pris en compte deux situations dans notre étude: le cas où les variables sont non-corrélées et le cas où les variables sont corrélées.

Dans les prochains paragraphes nous allons présenter la méthode de simulation utilisée, les résultats obtenus dans les deux cas de figure (variables corrélées et variables

indépendantes) et finalement nous allons analyser l'efficacité de la méthode de détection.

3.4.6.2.1 Méthode de simulation

Chaque variable a été générée selon une distribution normale respectivement de moyenne 5, 10 et 15 et d'écart-type $\sigma=1$. Nous avons simulé une région normale de fonctionnement de 50 échantillons (obs.1-50) pour les trois variables et ensuite nous avons simulé des dérèglages (sauts en échelon ou dérive en rampe, selon le cas) d'amplitude positive et négative de $1,5\sigma$ pour chaque variable:

- ✓ dérèglage d'amplitude négative pour la variable 1: obs.[51–100];
- ✓ dérèglage d'amplitude positive pour la variable 1: obs.[151–200];
- ✓ dérèglage d'amplitude négative pour la variable 2: obs.[251–300];
- ✓ dérèglage d'amplitude positive pour la variable 2: obs.[351–400];
- ✓ dérèglage d'amplitude négative pour la variable 3: obs.[451–500];
- ✓ dérèglage d'amplitude positive pour la variable 3: obs.[551–600];

On suppose, par la suite, que ces observations ont été enregistrées dans la base de données utilisateur et qu'elles ont participé à la formation des différentes classes de fonctionnement du processus. On suppose également que la bibliothèque de classes a été correctement constituée suite à l'étape d'analyse exploratoire des données dans le module d'analyse "off-line".

Nous avons précisé dans le paragraphe 3.2.4 que le taux d'erreur de la méthode de classement à l'aide de l'analyse discriminante est toujours surestimé si on utilise pour le test les observations qui ont participé au calcul des axes factoriels discriminants. Par conséquent, nous avons vérifié l'exactitude des résultats obtenus sur un ensemble simulé de données de test (obs.651–1250). Pour ces données, qui n'ont pas participé à la construction des axes factoriels discriminants, nous avons simulé des dérèglages (sauts en échelon ou dérives en rampe) ayant l'amplitude comprise entre $-3,5\sigma$ et $3,5\sigma$:

- ✓ dérèglage d'amplitude $-3,5\sigma$ pour la variable 1: obs.[651–670];
- ✓ dérèglage d'amplitude -3σ pour la variable 2: obs.[691–710];
- ✓ dérèglage d'amplitude $-2,5\sigma$ pour la variable 3: obs.[731–750];
- ✓ dérèglage d'amplitude -2σ pour la variable 1: obs.[771–790];
- ✓ dérèglage d'amplitude $-1,5\sigma$ pour la variable 2: obs.[811–830];
- ✓ dérèglage d'amplitude -1σ pour la variable 3: obs.[851–870];
- ✓ dérèglage d'amplitude $-0,5\sigma$ pour la variable 1: obs.[891–910];
- ✓ dérèglage d'amplitude $+0,5\sigma$ pour la variable 3: obs.[971–990];
- ✓ dérèglage d'amplitude $+1\sigma$ pour la variable 1: obs.[1111–1030];
- ✓ dérèglage d'amplitude $+1,5\sigma$ pour la variable 2: obs.[1051–1070];
- ✓ dérèglage d'amplitude $+2\sigma$ pour la variable 3: obs.[1091–1110];
- ✓ dérèglage d'amplitude $+2,5\sigma$ pour la variable 1: obs.[1131–1150];
- ✓ dérèglage d'amplitude $+3\sigma$ pour la variable 2: obs.[1171–1190];
- ✓ dérèglage d'amplitude $+3,5\sigma$ pour la variable 3: obs.[1211–1230].

3.4.6.2.2 Cas de trois variables non-corrélées

Dans ce paragraphe nous allons étudier la détection des dérèglages de type saut en échelon et dérive en rampe dans le cas de trois variables non-corrélées.

Cas 1: Dérèglages de type saut en escalier

Nous avons généré les données selon la méthode de simulation présentée dans le précédent paragraphe. Dans ce cas, le dérèglement est le saut en échelon. On présente dans la figure III.23.a les données avant le filtrage et dans les figures III.23.b, c et d les mêmes données après le filtrage avec des filtres passe-bas Butterworth d'ordre $L=2$ et ayant différentes fréquences de coupure ($FC=0.1$ Hz; $FC=0.06$ Hz et $FC=0.03$ Hz). Nous allons utiliser, par la suite dans cette étude, le filtre passe-bas avec une fréquence de coupure $FC=0.06$ Hz uniquement. Les figures III.23.b, c et d sont assez illustratives pour les transformations enregistrées dans la forme du signal suite à l'élimination de plus en plus prononcée des composantes spectrales hautes (les composantes du bruit en principal).

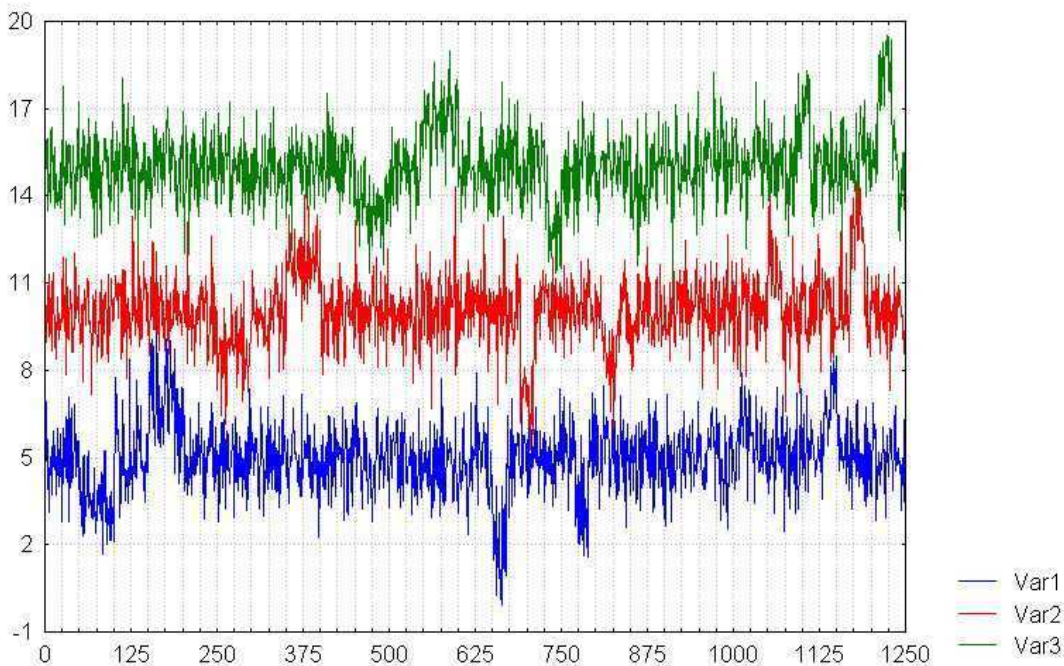


figure III.23.a – données avant le filtrage

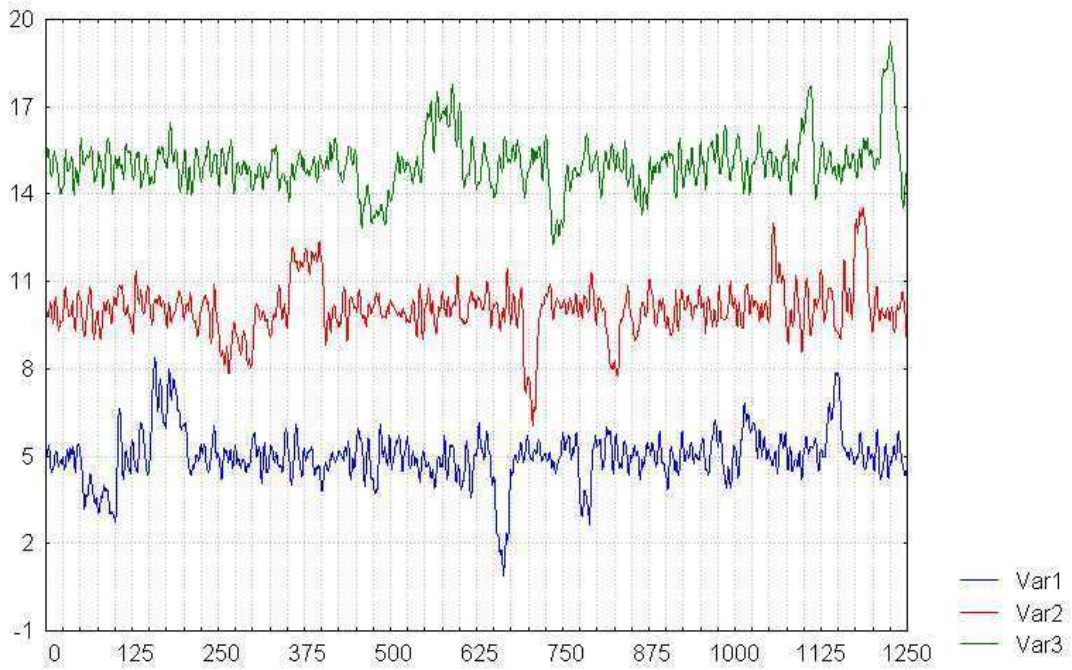


figure III.23.b – données filtrées
avec un filtre passe-bas Butterworth (L=2, FC = 0.1 Hz)

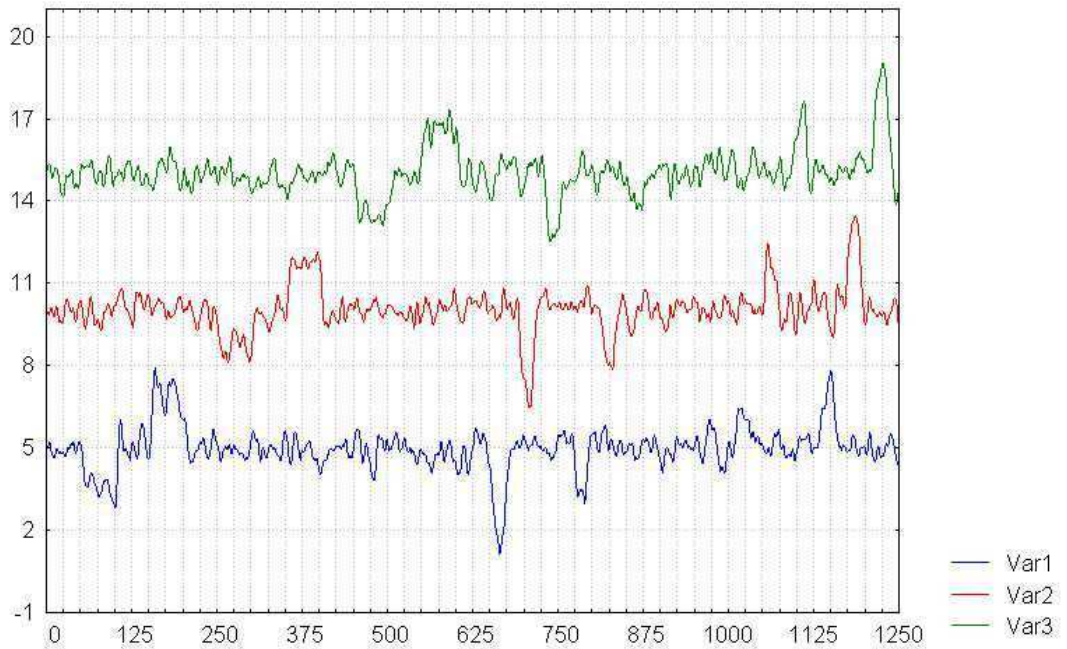


figure III.23.c – données filtrées
avec un filtre passe-bas Butterworth (L=2, FC = 0.06 Hz)

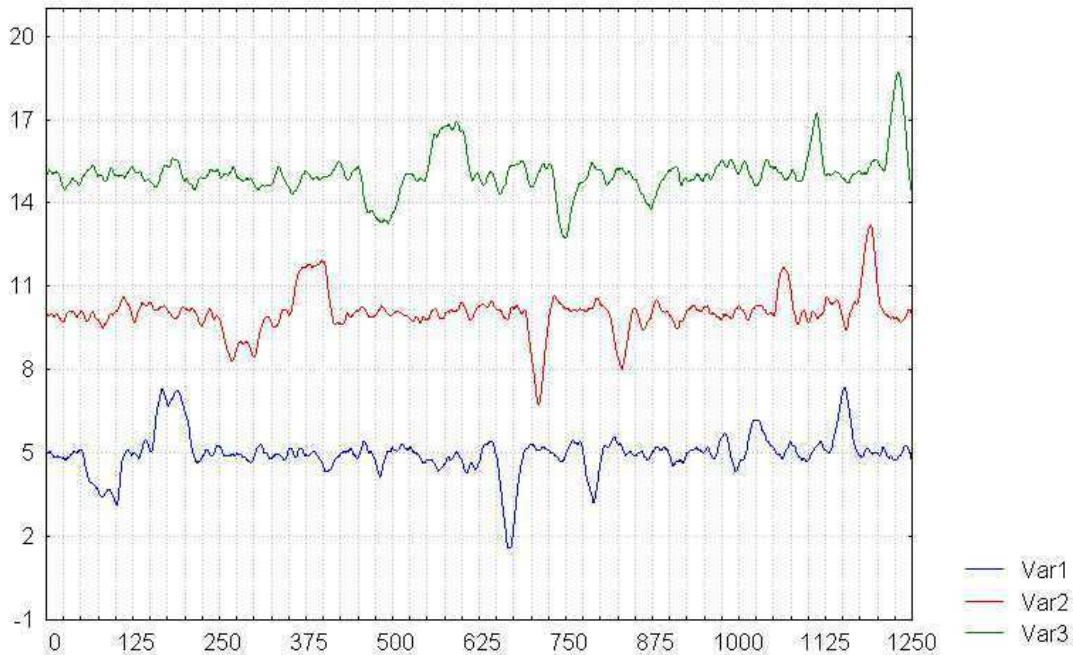


figure III.23.d – données filtrées
avec un filtre passe-bas Butterworth (L=2, FC = 0.03 Hz)

On a constaté, en utilisant le filtrage numérique, une amélioration notable de la discrimination entre les différentes classes de fonctionnement (au nombre de 7 dans ce cas: 6 classes de fonctionnement anormal du processus correspondant aux sauts positifs et négatifs de chaque variable et la classe de fonctionnement normal) suite à la réduction de la variabilité des variables (voir la situation présentée dans la figure III.23.c par rapport à la situation décrite dans la figure III.23.a). Par le filtrage, les classes deviennent plus agrégées et la distance entre les centres de gravité des classes augmente.

Nous avons calculé, pour 200 simulations, le taux moyen d'erreurs de classement. Suite à l'utilisation du filtrage numérique, le taux moyen d'erreurs de classement a été réduit à **5,42** sur l'échantillon participant à la construction des axes factoriels discriminants (obs.1-650) par rapport au taux moyen d'erreur de classement de **44,48** enregistré dans le cas des données initiales (non-filtrées).

L'amélioration de la discrimination entre les différentes classes de fonctionnement du processus suite à l'utilisation du filtrage numérique est également illustrée dans les figures III.24 et III.25. Dans la figure III.24, on a représenté la projection des données non-filtrées sur le plan principal de discrimination (formé par les deux premiers axes factoriels discriminants). Dans la figure III.25 on peut voir la projection sur le plan principal de discrimination des mêmes données mais après le filtrage.

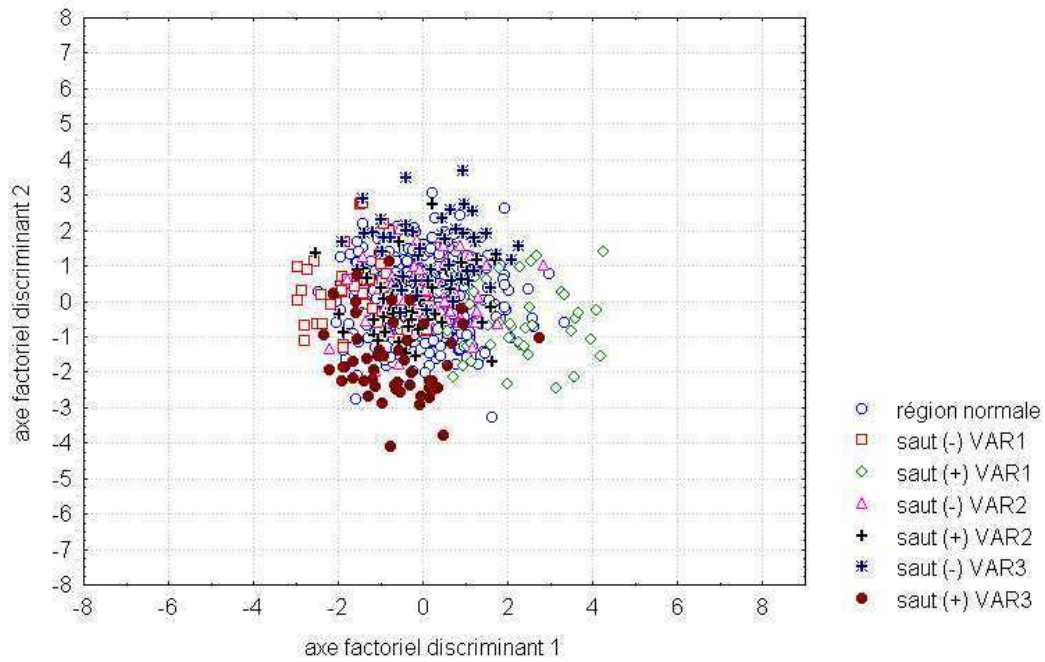


figure III.24 – observations non-filtrées dans le plan principal de discrimination

On peut constater, en comparant les deux figures, que le filtrage sépare mieux les 7 classes: la région normale de fonctionnement au centre et autour les six classes de dérèglement qui se délimitent presque parfaitement.

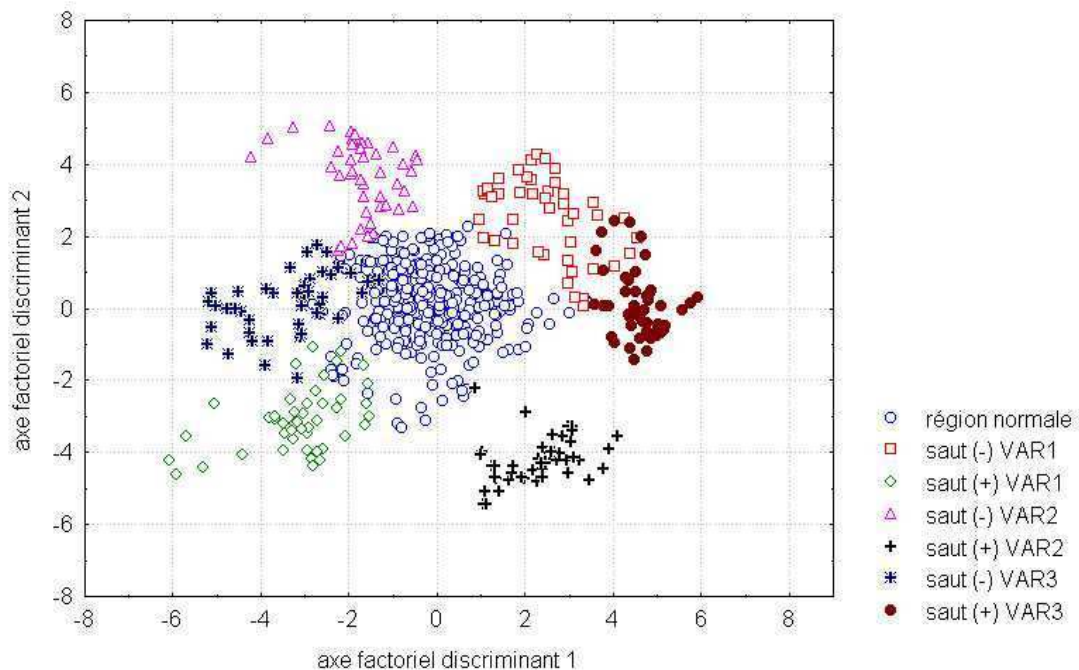


figure III.25 – observations filtrées dans le plan principal de discrimination

Nous avons analysé les performances de la méthode FNAD sur un échantillon de test (obs.651–1250) qui n'a pas participé à la construction des axes factoriels discriminants et où nous avons simulé des dérèglages de type saut en échelon d'amplitudes comprises entre $-3,5\sigma$ et $+3,5\sigma$.

Dans la figure III.26 nous pouvons constater la détection de ces dérèglages à l'aide de la carte de contrôle T^2 de Hotelling. Pour le calcul de la limite de contrôle on a utilisé un risque d'erreur de première espèce $\alpha=0,01$.

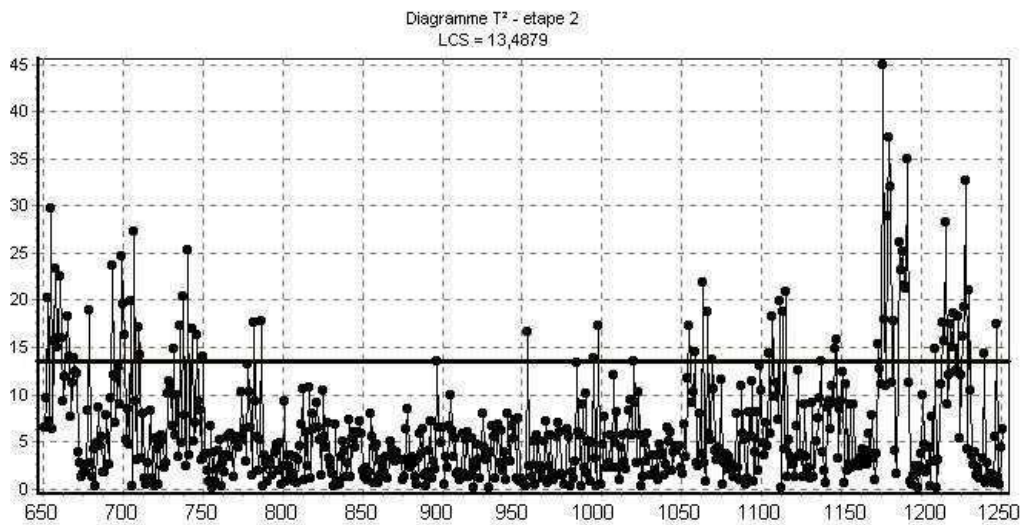


figure III.26 – détection des dérèglages sur la carte T^2 de Hotelling

Comme on peut le constater dans la fig.III.26 la carte T^2 de Hotelling détecte assez bien les différents dérèglages de la moyenne. Pourtant, il faut remarquer, que tous les points sur la carte T^2 ne dépassent pas la limite supérieure de contrôle quand on est dans une situation de dérèglage. Néanmoins, les quelques points qui dépassent la limite de contrôle supérieure nous suffiront pour détecter le dérèglage.

Dans le tableau III.5 on donne pour les mêmes observations les résultats obtenus à l'aide de la méthode FNAD. Dans ce tableau on a utilisé la codification suivante pour désigner les différentes classes de fonctionnement du processus:

- ✓ 1 - classe de fonctionnement normal;
- ✓ 110 - saut d'amplitude négative pour la variable 1;
- ✓ 111 - saut d'amplitude positive pour la variable 1;
- ✓ 120 - saut d'amplitude négative pour la variable 2;
- ✓ 121 - saut d'amplitude positive pour la variable 2;
- ✓ 130 - saut d'amplitude négative pour la variable 3;
- ✓ 131 - saut d'amplitude positive pour la variable 3;

Le premier dérèglage, correspondant au saut de $-3,5\sigma$ de la variable 1, est détecté sur le carte T^2 de Hotelling à partir de l'observation 652 qui dépasse la limite supérieure de contrôle. Donc, on sait qu'on se trouve en présence d'une cause assignable mais on ne sait pas quelle est la variable responsable de ce dérèglage. Cet inconvénient est caractéristique pour toutes les cartes de contrôle multivariées comme on l'a déjà mentionné dans le paragraphe 1.6.

Nous avons fait une étude comparative concernant les performances de détection des deux méthodes (T^2 de Hotelling et FNAD) et nous avons constaté que dans certaines cas, la carte de contrôle de Hotelling ne détecte pas le dérèglement mais que la méthode FNAD le détecte bien. C'est particulièrement le cas dans cet exemple, avec le dérèglement de $-1,5\sigma$ simulé à partir de l'observation 811 jusqu'à l'observation 830 qui n'a pas été détecté sur la carte T^2 de Hotelling et qui est détecté à partir de l'observation 814 à l'aide de la méthode FNAD.

Efficacité de la méthode

Nous avons calculé, pour 250 simulations sur des données de test (qui n'ont pas participé à la construction des axes discriminants), le pourcentage de détection des dérèglements et la POM_1 (Période Opérationnelle Moyenne) de la méthode FNAD. Les résultats sont présentés dans le tableau III.6:

tableau III.6 - performances de la méthode FNAD

Amplitude du saut	$-3,5\sigma$	-3σ	$-2,5\sigma$	-2σ	$-1,5\sigma$	-1σ	$-0,5\sigma$	$+0,5\sigma$	$+1\sigma$	$+1,5\sigma$	$+2\sigma$	$+2,5\sigma$	$+3\sigma$	$+3,5\sigma$
Détection (%)	100	100	100	100	100	95,6	47,2	40,8	98	100	100	100	100	100
POM_1 (écart-type)	3,4 (0,7)	3,69 (0,82)	4,05 (0,95)	4,54 (1,24)	5,7 (2,11)	9,19 (4,47)	11,81 (5,38)	11,51 (5,17)	8,54 (4,22)	5,82 (2,07)	4,72 (1,18)	3,92 (0,91)	3,62 (0,82)	3,39 (0,83)

En ce qui concerne la détection des dérèglements, on peut affirmer que la méthode donne de très bons résultats. Pour des dérèglements dont l'amplitude varie de $-3,5\sigma$ à $-1,5\sigma$ et de $1,5\sigma$ à $3,5\sigma$ le pourcentage de détection est de 100. La méthode présente aussi des bons résultats pour des dérèglements faibles, car pour les dérèglements de -1σ et $+1\sigma$ le pourcentage de détection est élevé (respectivement 95,6% et 98%). La POM_1 de la méthode présente de très bonnes valeurs pour des dérèglements d'amplitudes supérieures à $1,5\sigma$. Pour les dérèglements faibles la détection s'effectue avec un retard d'environ 9 observations par rapport au moment de l'apparition du dérèglement.

Cas 2: Dérives en rampe

Les dérèglements lents (de type dérive rampe) sont plus fréquents en pratique, c'est pourquoi dans le présent paragraphe nous avons étudié la détection de ce type de dérèglements à l'aide de la méthode FNAD. Nous avons utilisé la même démarche que dans le paragraphe précédent, la seule différence étant que cette fois-ci nous avons testé les performances de la méthode sur des données contenant des sauts en rampe au lieu des sauts en escalier. Les résultats, concernant le pourcentage de détection et la POM_1 , pour un ensemble de 250 simulations sont présentés dans le tableau III.7:

tableau III.7 - performances de la méthode FNAD

Amplitude du saut	$-3,5\sigma$	-3σ	$-2,5\sigma$	-2σ	$-1,5\sigma$	-1σ	$-0,5\sigma$	$+0,5\sigma$	$+1\sigma$	$+1,5\sigma$	$+2\sigma$	$+2,5\sigma$	$+3\sigma$	$+3,5\sigma$
Détection (%)	100	100	100	100	89,2	67,2	23,6	26,4	60,4	92,8	99,6	100	100	100
POM_1 (écart-type)	9,18 (2,44)	10,28 (2,39)	11,3 (2,96)	12,96 (3,6)	14,95 (3,75)	15,54 (4,59)	15,49 (4,68)	15,95 (4,59)	15,29 (4,64)	14,51 (4,26)	12,95 (3,73)	11,43 (2,93)	10,48 (2,52)	9,43 (2,32)

La capacité de détection de la méthode FNAD reste bonne pour des dérèglements ayant des amplitudes supérieures à $1,5\sigma$. Pour des dérèglements de faibles amplitudes la

méthode perd de son efficacité. En ce qui concerne la POM_1 , les valeurs sont relativement élevées, mais il ne faut pas perdre de vue que le dérèglement est en rampe et qu'il est plus difficile de le détecter car il ne se manifeste pas instantanément.

3.4.6.2.3 Cas des trois variables corrélées

Le plus souvent, en pratique, les variables de processus ne sont pas indépendantes mais corrélées. Dans ce paragraphe nous analyserons l'effet de la corrélation entre les variables sur les performances de la méthode FNAD.

Dans ce but, on considère dans une première étape, la situation décrite par Hawkins [47], où les variables sont corrélées mais le dérèglement de l'une n'influence pas les autres. Hawkins donne, dans ce sens, l'exemple du contrôle statistique d'un processus de raffinage de platine. Nous avons généré des données corrélées selon une matrice de variance-covariance $\Sigma = 0.2 \times \mathbf{I} + 0.8 \times \mathbf{1} \times \mathbf{1}^T$, où \mathbf{I} est la matrice identité de dimension 3×3 et $\mathbf{1}$ est un vecteur de dimension 3×1 ayant tous les éléments égaux à 1. La forte corrélation entre les variables, change la disposition des points dans l'espace \mathbb{R}^3 . Ainsi, au lieu d'avoir des points disposés sous forme d'un nuage à peu près sphérique, comme c'était le cas pour des données non-corrélées, la forte corrélation entre les variables allonge l'allure du nuage le transformant en un ellipsoïde semblable à un ballon de rugby. Plus la corrélation est importante et plus l'ellipsoïde est allongé. Il est évident que dans ce cas, un éventuel dérèglement est plus facile à détecter, car les points correspondant au dérèglement vont se distinguer clairement des autres points correspondant au fonctionnement normal du processus (voir les points en carré sur la figure III.27 où nous avons simulé un dérèglement de type saut échelon d'amplitude 1σ pour la variable 1).

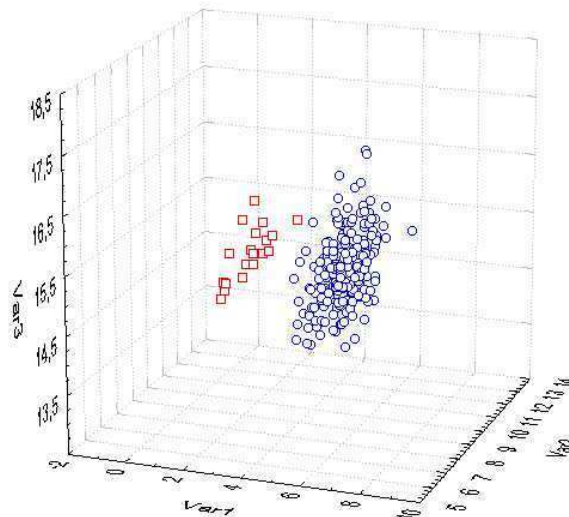


figure III.27 - disposition des points dans l'espace \mathbb{R}^3 (cas des variables corrélées)

Le filtrage numérique va accentuer encore davantage la séparation de six zones de dérèglement par rapport à la zone de fonctionnement normal. Dans la figure III.28 on donne pour le même ensemble des données corrélées, la projection des points sur le plan des deux premiers axes factoriels discriminants. En comparant les figures III.28 et III.25 on voit bien que la forte corrélation entre les données, loin d'être un inconvénient

comme dans la majorité des méthodes de contrôle statistique multivariées, nous aide à mieux discriminer les sept classes.

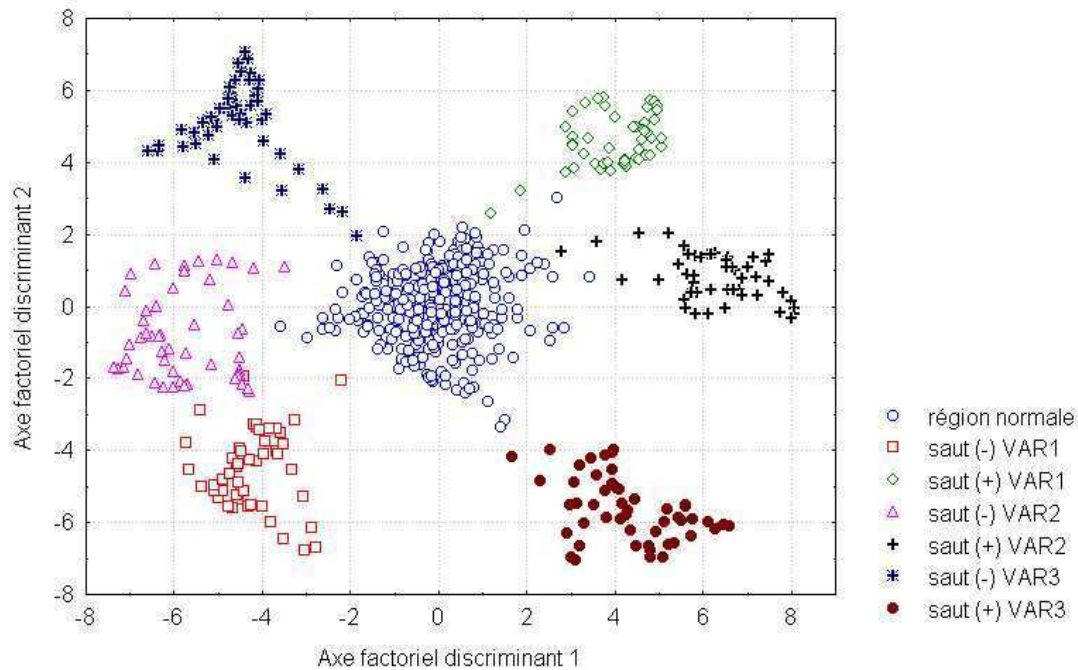


figure III.28 - projection des données corrélées sur le plan discriminant principal après le filtrage numérique

Ainsi, en étudiant le taux moyen d'erreurs de classement, pour les données qui ont participé à la construction des axes factoriels discriminants (*obs1-650*), nous avons obtenu pour les 250 simulations la valeur de **1,2**.

Dans le tableau III.8 on donne, pour 250 simulations, les performances de la méthode FNAD dans le cas des variables corrélées:

tableau III.8 - performances de la méthode FNAD

Amplitude du saut	-3,5σ	-3σ	-2,5σ	-2σ	-1,5σ	-1σ	-0,5σ	+0,5σ	+1σ	+1,5σ	+2σ	+2,5σ	+3σ	+3,5σ
Détection (%)	100	100	100	100	100	100	31,6	22	100	100	100	100	100	100
POM ₁ (écart-type)	3,14 (0,46)	3,38 (0,6)	3,67 (0,52)	4,21 (0,65)	5,12 (1,1)	7,11 (2,21)	12,32 (4,76)	13,05 (4,18)	7,46 (2,52)	5,09 (1,11)	4,15 (0,59)	3,78 (0,55)	3,43 (0,54)	3,13 (0,37)

On constate, en regardant le tableau III.8, que par rapport à la situation précédente (*données non-corrélées* - tableau III.6), la détection des dérèglages s'est améliorée pour les sauts de faibles amplitudes (-1σ,+1σ). La POM₁ de la méthode s'est aussi améliorée mais pas d'une manière très significative. En tout cas, on peut conclure que les performances de la méthode FNAD ne sont pas diminuées quand les données sont corrélées, mais qu'au contraire elles sont sensiblement supérieures à celles enregistrées dans la situation des données non-corrélées.

Toutes les méthodes de contrôle multivariées basées sur la statistique T² de Hotelling, utilisent dans le calcul de la distance Mahalanobis l'inverse de la matrice de variance-covariance Σ. Si les variables sont fortement corrélées, cette matrice est

approximativement singulière et donc non-inversible. Nous avons vu précédemment que la méthode FNAD présente des performances très bonnes quand les données sont corrélées, cela étant un avantage majeur par rapport aux méthodes de type Hotelling.

Vue la corrélation très forte entre les variables il nous semble intéressant d'étudier aussi la situation, très probable, dans laquelle plusieurs variables changent dans le même temps. Ainsi, on a testé les performances de la méthode dans la détection des dérèglages dans toutes les directions (c'est à dire pour toutes les combinaisons possibles des dérèglages pour les trois variables). Pour cela, nous avons simulé des dérèglages d'amplitude $1,5\sigma$. Pour trois variables, on a 27 classes de fonctionnement: la classe de fonctionnement normal et 26 classes correspondant à toutes les combinaisons de dérèglages possibles. Nous avons utilisé un code de trois chiffres (**XXX**) pour codifier les classes: le premier chiffre dans le code est pour désigner l'état de la première variable, le deuxième chiffre désigne l'état de la deuxième variable et, finalement, le troisième chiffre codifie l'état de la troisième variable. La signification des chiffres est la suivante:

- ✓ 0 - la variable respective n'a pas changé;
- ✓ 1 - la variable respective a subi un dérèglage d'amplitude négative;
- ✓ 2 - la variable respective a subi un dérèglage d'amplitude positive;

Ainsi, par exemple, le code 010 signifie un dérèglage d'amplitude négative pour la deuxième variable, alors que le code 121 signifie un dérèglage d'amplitude négative pour la première et la troisième variable et d'amplitude positive pour la deuxième variable.

Pour la même matrice de variance-covariance que dans le cas précédent, nous avons étudié les performances de la méthode dans la détection des différents types de dérèglages. Les pourcentages de détection et la POM_1 de la méthode sont présentés dans le tableau III.9. Pour des dérèglages ayant une amplitude supérieure à $1,5\sigma$ la méthode détecte bien dans toutes les directions. La POM_1 de la méthode reste aussi dans les mêmes limites que pour le cas précédent. On peut conclure que la méthode FNAD détecte bien les dérèglages simultanés de plusieurs variables sans favoriser une certaine direction dans l'espace multidimensionnel. De ce point de vue, on peut dire que la méthode FNAD est plus performante que la méthode de détection reposant sur l'analyse en composante principale (ACP).

tableau III.9 - performances de la méthode FNAD

Amplitude Saut	1 sigma		1,5 sigma		2 sigma		2,5 sigma		3 sigma	
	Déréglage	det (%)	POM ₁ (ec-t)	det (%)	POM ₁ (ec-t)	det (%)	POM ₁ (ec-t)	det (%)	POM ₁ (ec-t)	det (%)
001	92,40	6,66 (1,81)	100	5,14 (1,34)	100	4,12 (0,80)	100	3,80 (1,19)	100	3,45 (1,24)
002	93,20	6,26 (1,72)	100	4,76 (1,34)	100	4,01 (1,34)	100	3,56 (1,31)	100	3,26 (1,31)
010	90,40	6,63 (1,89)	99,6	5,02 (1,35)	100	4,10 (0,97)	100	3,78 (1,07)	100	3,64 (1,70)
011	73,60	8,20 (1,43)	98,8	5,91 (1,00)	100	4,85 (0,68)	100	4,19 (0,58)	100	3,86 (0,47)
012	94,00	6,87 (1,67)	100	5,04 (1,14)	100	4,25 (0,99)	100	3,90 (1,02)	99,6	3,72 (1,48)
020	89,20	6,37 (1,88)	100	4,94 (1,30)	100	4,08 (1,21)	99,6	3,55 (1,37)	99,6	3,18 (1,06)
021	87,20	7,17 (1,85)	99,6	5,37 (1,27)	100	4,56 (1,16)	100	4,14 (1,19)	99,6	3,90 (1,36)
022	83,60	7,62 (1,61)	100	5,57 (1,09)	100	4,60 (0,78)	100	4,02 (0,49)	100	3,66 (0,50)
100	84,00	7,06 (1,93)	99,2	5,34 (1,48)	99,6	4,39 (1,20)	100	3,91 (1,32)	100	3,63 (1,27)
101	77,60	7,52 (1,77)	98,8	5,77 (1,29)	100	4,79 (1,03)	100	4,24 (1,05)	100	3,82 (0,61)
102	52,00	8,38 (1,43)	100	6,10 (1,02)	100	4,93 (0,64)	100	4,35 (0,56)	100	4,05 (0,74)
110	37,60	8,87 (1,17)	100	6,56 (0,86)	99,6	5,16 (0,57)	100	4,63 (0,52)	100	4,15 (0,72)
111	30,40	9,25 (1,00)	100	6,75 (0,75)	100	5,32 (0,50)	100	4,74 (0,45)	100	4,19 (0,39)
112	35,60	9,08 (1,09)	99,6	6,60 (0,72)	100	5,31 (0,54)	100	4,65 (0,48)	100	4,16 (0,37)
120	82,00	6,95 (1,93)	99,2	5,54 (1,55)	100	4,61 (0,99)	100	4,23 (1,23)	99,6	3,84 (1,15)
121	84,00	6,62 (2,48)	99,6	5,13 (1,69)	100	4,28 (1,02)	100	3,99 (0,96)	100	3,61 (0,73)
122	69,20	8,09 (1,42)	100	5,86 (0,95)	100	4,80 (0,59)	100	4,20 (0,52)	100	3,86 (0,54)
200	84,40	6,93 (1,76)	100	5,21 (1,14)	100	4,44 (1,42)	100	3,82 (0,96)	100	3,42 (1,08)
201	58,80	8,23 (1,35)	100	6,13 (0,95)	100	5,04 (0,93)	100	4,41 (0,78)	100	4,02 (0,34)
202	79,60	7,52 (1,83)	100	5,93 (1,40)	100	4,74 (0,81)	100	4,33 (1,16)	100	4,07 (1,50)
210	85,60	6,71 (1,95)	99,6	5,31 (1,63)	100	4,35 (1,05)	100	3,92 (0,81)	100	3,58 (0,84)
211	58,80	8,52 (1,51)	99,2	6,21 (1,02)	100	5,01 (0,61)	100	4,43 (0,53)	100	4,03 (0,31)
212	87,20	6,73 (2,16)	99,2	5,09 (1,71)	100	4,32 (1,26)	100	3,73 (0,99)	100	3,57 (0,87)
220	40,80	8,95 (1,15)	100	6,40 (0,76)	100	5,10 (0,60)	100	4,37 (0,55)	100	4,06 (0,71)
221	20,40	9,20 (1,24)	100	6,81 (0,84)	100	5,38 (0,54)	100	4,79 (0,43)	100	4,17 (0,37)
222	38,80	8,97 (1,07)	100	6,41 (0,69)	100	5,21 (0,46)	100	4,62 (0,49)	100	4,10 (0,31)

3.5 Application industrielle

3.5.1 Introduction

Dans ce paragraphe nous analyserons un exemple pratique d'application de la méthode FNAD. Nous considérons, dans ce but, les deux premières sections d'un réacteur utilisé pour la fabrication à haute pression de polyéthylène de faible densité.

Le polyéthylène est un des polymères les plus répandus fabriqués actuellement. En 1990 la production mondiale de polyéthylène a été estimée à 25×10^6 tonnes par an (Kiparissides et al. [62]): 65% étant du polyéthylène de faible densité obtenu dans des réacteurs à haute pression et 35% étant du polyéthylène de haute densité produit dans des réacteurs à basse pression. La densité et le taux de ramification sont deux des caractéristiques physiques et moléculaires du polyéthylène les plus importantes. Si la densité est comprise dans l'intervalle $[0.91 - 0.925 \text{ g/cm}^3]$ le polyéthylène est considéré comme étant de faible densité; si la densité est dans l'intervalle $[0.926 - 0.94 \text{ g/cm}^3]$ le polyéthylène est considéré comme étant de moyenne densité et finalement si la densité est dans l'intervalle $[0.941 - 0.965 \text{ g/cm}^3]$ le polyéthylène est considéré comme étant de haute densité. La densité du polyéthylène est fonction du nombre de ramifications courtes dans sa composition moléculaire. Plus le nombre de ramification courtes est réduit et plus la densité du polyéthylène est haute.

Il faut préciser ici que le type de ramifications (courtes ou longues) dans la structure moléculaire du polyéthylène, la fonctionnalité, la forme, et la distribution des ramifications sont fortement liés au processus de polymérisation et aux conditions opérationnelles du réacteur. Par conséquent, un objectif majeur pour les ingénieurs et les spécialistes du processus est l'identification de l'influence que peuvent avoir les conditions opérationnelles du réacteur sur la "qualité" du polymère. Selon Kiparissides et al. [65], la production "intelligente" du polyéthylène doit intégrer d'une façon harmonieuse la modélisation du processus, les capteurs performants issus d'une technologie moderne et les approches statistiques de contrôle du processus.

3.5.2 Le processus de fabrication du polyéthylène

La fabrication du polyéthylène de faible densité est réalisée en présence de faibles quantités de catalyseurs (peroxydes, composantes à base d'azote, etc.) dans des réacteurs, à haute pression (200-350 Mpa.) et à haute température (150-300 °C). Dans ces conditions, l'éthylène se trouve sous forme liquide à l'intérieur du réacteur.

Le réacteur, dont le schéma réduit est présenté dans la figure III.29, est constitué d'un tube métallique ayant un grand rapport longueur/diamètre (la longueur du réacteur est comprise entre 600 et 1200 m et son diamètre intérieur ne dépasse pas 70-80 mm). L'éthylène, le solvant et le catalyseur sont injectés dans la première section du réacteur. Des injections supplémentaires d'éthylène, de solvant et de catalyseur sont effectuées entre les deux sections. Le schéma complet du réacteur, qui n'est pas présenté ici, contient trois modules: le module de compression, le réacteur proprement dit et le module de séparation des constituants. Du point de vue thermique le réacteur peut être divisé également en trois modules: le module de préchauffement, le module de réaction et le module de refroidissement. Un réacteur industriel peut disposer de 3 à 5 modules

de réaction et de plusieurs modules de refroidissement. Les conditions de réaction (la température, la pression, la concentration du catalyseur, etc.) peuvent être réglées indépendamment dans chaque module afin d'obtenir des polymères ayant une large gamme de masses moléculaires.

La réaction chimique de polymérisation de l'éthylène est très exotherme et, par conséquent, chaque section dispose d'un revêtement circulaire dans lequel circule un liquide de refroidissement qui contribue à l'évacuation de la moitié de la chaleur dégagée par la réaction. La pression à l'intérieur du réacteur est périodiquement diminuée, pour une très courte période de temps, à l'aide d'une valve située à la fin du tube, afin de diminuer les dépôts de polymérisation sur les parois du réacteur, dépôts qui empêchent l'évacuation de la chaleur accumulée à l'intérieur du réacteur.

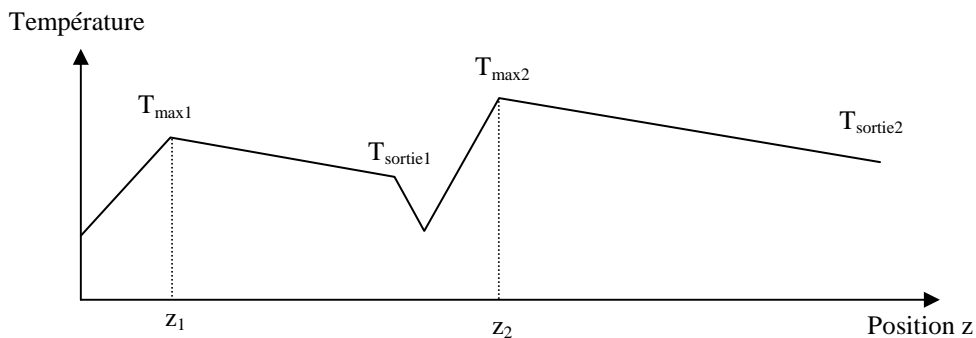
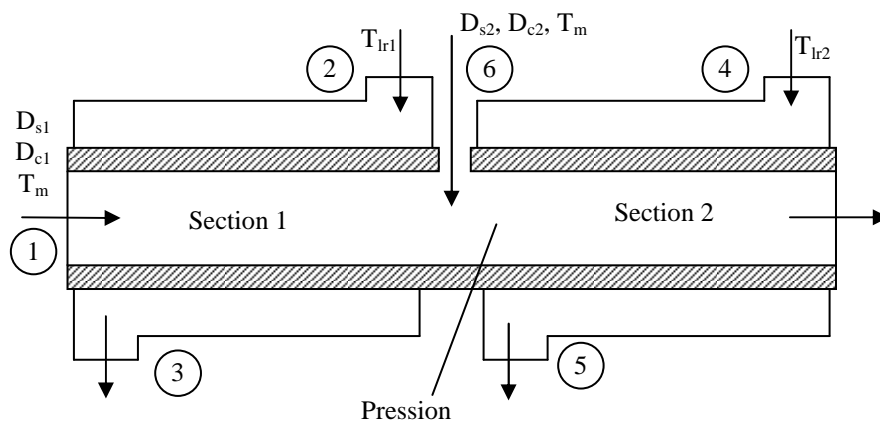


figure III.29 - schéma et courbe de température du réacteur

Légende:

1- l'entrée dans la première section du réacteur; 2 et 4 - les entrées du liquide de refroidissement; 3 et 5 - les sorties du liquide de refroidissement; 6 - l'entrée dans la deuxième section du réacteur.

3.5.3 Modélisation mathématique du processus de fabrication

Kiparissides et al. [62] proposent un modèle mathématique pour le processus de fabrication du polyéthylène. En utilisant la méthode des moments pour résoudre un système d'équations différentielles, ils déduisent les équations de conservation de la masse totale, de l'énergie et du moment. Les auteurs donnent, également, les équations du développement de la masse moléculaire du polymère et les équations caractérisant le changement de composition dans les deux sections du réacteur. En plus, le modèle proposé est basé sur des équations décrivant la variation des grandeurs physiques, thermodynamiques et liées au déplacement dans le réacteur du mélange des constituants chimiques, en fonction de la température, de la pression et de la composition du polymère.

A l'aide du modèle mathématique ainsi construit plusieurs courbes d'évolution pour les paramètres de processus comme, par exemple, la courbe de la température à l'intérieur du réacteur (figure III.29) ont été obtenues. L'explication du phénomène physique associé à cette courbe est très simple: après l'injection du catalyseur (le point 1 sur la figure III.29) la réaction chimique se consomme rapidement en dégageant une grande quantité de chaleur; la température maximale $T_{\max 1}$ sur la courbe de la température (figure III.29) correspond à la fin de la réaction (l'épuisement du catalyseur); le mélange obtenu est, par la suite, refroidi à l'aide du liquide de refroidissement qui coule dans le revêtement circulaire entourant le réacteur et donc la température diminue lentement; à la fin de la première section le mélange est refroidi par l'addition d'une quantité supplémentaire de monomère et le tout est passé dans la deuxième section; dans la deuxième section le processus chimique se déroule de la même façon.

En analysant, à l'aide du modèle, l'évolution de différents paramètres (qualitatifs ou de processus), Kiparissides et al. [62] ont tiré les conclusions suivantes:

- ✓ l'augmentation de la concentration du catalyseur a pour effet:
 - o l'augmentation de la température à l'intérieur du réacteur (T);
 - o le déplacement sur l'axe z de la position du maximum de la température (figure III.29) vers l'entrée du réacteur (vers la gauche);
 - o l'augmentation du taux de conversion du monomère (Conv.);
 - o l'augmentation de la masse moléculaire moyenne en masse du monomère (Mw);
 - o l'augmentation du taux de ramifications longues (TRL) et du taux de ramifications courtes (TRC);
 - o la diminution de la masse moléculaire moyenne en nombre du monomère (Mn);
 - o la diminution du coefficient de conductivité thermique du mélange;
 - o l'augmentation de la viscosité du mélange (due à l'augmentation de la conversion du monomère);
- ✓ l'augmentation modérée de la concentration du solvant a pour effet:
 - o la diminution importante de la masse moléculaire moyenne en masse du monomère (Mw);
 - o la diminution importante de la masse moléculaire moyenne en nombre du monomère (Mn);
- ✓ l'augmentation de la pression à l'intérieur du réacteur a pour effet:
 - o l'augmentation de la température (T) dans le réacteur;
 - o l'augmentation du taux de conversion du monomère (Conv);

- ✓ l'augmentation du débit du mélange dans le réacteur a comme effet:
 - o l'augmentation de la résistance thermique des parois du réacteur à cause de l'augmentation des dépôts de polymère;
 - o l'augmentation de la température à l'intérieur du réacteur;

3.5.4 Choix des variables suivies

Ce processus de fabrication du polyéthylène a été également analysé par Kourti et MacGregor [65] à l'aide des méthodes de projection (ACP, PSL). Ils proposent le suivi de deux catégories de variables pour la maîtrise du processus:

1. les variables de processus:

- ✓ $T_{\max 1}$ et $T_{\max 2}$ – la température maximale du mélange dans le réacteur;
- ✓ $T_{\text{sortie}1}$ et $T_{\text{sortie}2}$ – la température du mélange à la sortie du réacteur;
- ✓ $T_{\text{lr}1}$ et $T_{\text{lr}2}$ – la température du liquide de refroidissement;
- ✓ z_1 et z_2 – la position de la $T_{\max 1}$ et de la $T_{\max 2}$ dans le réacteur;
- ✓ D_{c1} et D_{c2} – le débit du catalyseur;
- ✓ D_{s1} et D_{s2} – le débit du solvant;
- ✓ T_m – la température du mélange introduit dans le réacteur;
- ✓ Pression – la pression dans le réacteur;

Remarque:

Les indices 1 et 2 utilisés ci-dessus désignent l'appartenance des variables à la section 1 ou à la section 2 du réacteur.

2. les caractéristiques qualité du produit:

- ✓ le taux de conversion du monomère (Conv);
- ✓ la masse moléculaire moyenne en masse du monomère (Mw);
- ✓ la masse moléculaire moyenne en nombre du monomère (Mn);
- ✓ le taux de ramifications longues (TRL) /1000 atomes de C;
- ✓ le taux de ramifications courtes (TRC) /1000 atomes de C;

Dans cette étude, pour simplifier, nous allons seulement prendre en compte le suivi des variables de processus, car le suivi des caractéristiques de qualité du produit peut se faire de la même manière. De plus, très souvent en pratique, les caractéristiques de qualité du produit ne sont pas disponibles à chaque instant, étant mesurées moins souvent que les variables de processus (Kourti et MacGregor [65]). Ce manque de données qualitatives vient du fait que ces caractéristiques sont analysées dans le laboratoire et cette analyse peut nécessiter souvent un temps assez long avant de fournir les résultats. Par conséquent, de nombreux produits défectueux seront fabriqués avant que les résultats de laboratoire soient enfin disponibles. Dans cet exemple, aucune caractéristique qualitative du produit n'est disponible en temps réel. Par contre, plusieurs variables de processus, comme la température du liquide de refroidissement, le débit du solvant, le débit du catalyseur et la courbe d'évolution de la température à l'intérieur du réacteur sont disponibles très fréquemment. L'apparition d'une cause spéciale qui affecte les caractéristiques qualitatives du produit pourra, très probablement, être détectée avec succès dans l'espace des variables de processus. A part l'aspect déjà mentionné, il y a

également d'autres raisons pour un suivi à l'aide des variables de processus: premièrement, très souvent en pratique, peu de caractéristiques de qualité sont mesurées et ne suffisent pas pour définir entièrement la qualité du produit final; deuxièmement, même si les caractéristiques de qualité étaient disponibles plus fréquemment, permettant ainsi l'identification en temps utile de la caractéristique qui a changé, il est difficile pourtant de préciser ce qui s'est passé dans le processus et donc de prendre les actions correctives nécessaires. Par le suivi des variables de processus on peut, dans la plupart des cas, trouver plus facilement l'origine de la cause assignable.

3.5.5 Création de la base de données

Nous avons simulé dans cette base de données l'apparition des "événements" suivants:

- ✓ dépôts de polymère sur les parois du réacteur;
- ✓ augmentation du taux d'impuretés dans le mélange;
- ✓ augmentation de la concentration du catalyseur dans la première section et dans la deuxième section;
- ✓ augmentation de la concentration du solvant dans la première section et dans la deuxième section;
- ✓ augmentation de la pression dans le réacteur;

Ces "événements" (ou causes assignables) se produisent très souvent en pratique et conditionnent la qualité finale du polyéthylène. Il est donc nécessaire de les détecter et les éliminer afin de réduire le nombre de produits non-conformes et le coût associé à la production.

On peut constater que ces causes assignables peuvent être groupées en deux classes:

- ✓ causes assignables qui ne peuvent pas être mesurées directement (le dépôt de polymère sur les parois et l'augmentation du taux d'impuretés dans le mélange) mais qui influencent un certain nombre de variables de processus mesurées fréquemment;
- ✓ causes assignables qui peuvent être mesurées directement (la pression, le débit) et qui, éventuellement, peuvent influencer les autres variables de processus (c'est notamment le cas de la pression et du débit du catalyseur dans cet exemple);

Nous avons résumé dans le tableau III.10 l'effet de chacune de ces causes assignables sur les variables de processus*.

Il faut préciser ici que dans cet exemple nous avons considéré que toutes les données disponibles dans la base de données ont été obtenues dans des conditions d'équilibre.

* Selon l'étude et les conclusions de Kiparissides et al. [62] et Kourti et MacGregor [65]

tableau III.10

Causes assignables	Effets sur les variables de processus
dépôts de polymère sur les parois du réacteur	$T_{\text{sortie1}} (\uparrow); T_{\text{sortie2}} (\uparrow);$
taux d'impuretés dans le mélange (\uparrow)	$T_{\text{sortie1}} (\downarrow); T_{\text{sortie2}} (\downarrow); T_{\text{max1}} (\downarrow); T_{\text{max2}} (\downarrow);$ $z_1 (\uparrow) \text{ et } z_2 (\uparrow);$
$D_{c1} (\uparrow)$	$T_{\text{max1}} (\uparrow); T_{\text{sortie1}} (\uparrow); z_1 (\downarrow);$
$D_{c2} (\uparrow)$	$T_{\text{max2}} (\uparrow); T_{\text{sortie2}} (\uparrow); z_2 (\downarrow);$
$D_{s1} (\uparrow)$ mais modérée	pas d'influence sur les autres variables de processus;
$D_{s2} (\uparrow)$ mais modérée	pas d'influence sur les autres variables de processus;
Pression: $P (\uparrow)$	$T_{\text{max1}} (\uparrow); T_{\text{sortie1}} (\uparrow); z_1 (\downarrow); T_{\text{max2}} (\uparrow); T_{\text{sortie2}} (\uparrow); z_2 (\downarrow);$

La région de fonctionnement normal du processus est caractérisée par les valeurs suivantes pour les paramètres de processus:

- ✓ la température du mélange à l'entrée dans le réacteur: 192 °C;
- ✓ la température du liquide de refroidissement: 127 °C;
- ✓ le débit du liquide de refroidissement: 50 kg/s;
- ✓ la pression nominale à l'intérieur du réacteur: 250 Mpa;
- ✓ la concentration du catalyseur pour les deux sections: 4E-5 kg mol/m³;
- ✓ la température T_{max1} : 315 °C;
- ✓ la température T_{max2} : 310 °C;
- ✓ la température T_{sortie1} : 290 °C;
- ✓ la température T_{sortie2} : 295 °C;
- ✓ la position relative z_1 : 0.15;
- ✓ la position relative z_2 : 0.7;

Dans la base de données utilisateur nous avons simulé 50 observations pour chaque classe de fonctionnement anormal du processus et 400 observations pour la classe de fonctionnement normal du processus. Nous avons considéré que certaines variables sont indépendantes (D_{s1} , D_{s2}) et que certaines variables sont corrélées (D_{c1} , D_{c2} , T_{max1} , T_{sortie1} , T_{max2} , T_{sortie2} , z_1 , z_2 et P). Dans le tableau III.11 on donne la matrice de corrélation des variables. On peut remarquer, en gras dans ce tableau, les fortes corrélations entre les variables de processus.

tableau III.11

	P	D_{s1}	D_{s2}	D_{c1}	D_{c2}	T_{max1}	T_{sortie1}	z_1	T_{max2}	T_{sortie2}	z_2
P		-0,02	0,00	0,66	0,61	0,71	0,63	-0,62	0,69	0,62	-0,56
D_{s1}	-0,02		0,00	-0,03	-0,03	-0,02	-0,03	0,01	-0,02	-0,02	0,01
D_{s2}	0,00	0,00		-0,05	0,05	-0,06	-0,05	0,04	0,06	0,02	-0,04
D_{c1}	0,66	-0,03	-0,05		0,02	0,90	0,82	-0,74	0,03	0,04	0,00
D_{c2}	0,61	-0,03	0,05	0,02		-0,02	-0,03	-0,03	0,90	0,82	-0,74
T_{max1}	0,71	-0,02	-0,06	0,90	-0,02		0,91	-0,84	-0,01	-0,01	0,03
T_{sortie1}	0,63	-0,03	-0,05	0,82	-0,03	0,91		-0,76	-0,02	-0,02	0,02
z_1	-0,62	0,01	0,04	-0,74	-0,03	-0,84	-0,76		-0,02	-0,02	0,03
T_{max2}	0,69	-0,02	0,06	0,03	0,90	-0,01	-0,02	-0,02		0,91	-0,83
T_{sortie2}	0,62	-0,02	0,02	0,04	0,82	-0,01	-0,02	-0,02	0,91		-0,78
z_2	-0,56	0,01	-0,04	0,00	-0,74	0,03	0,02	0,03	-0,83	-0,78	

3.5.6 Module d'analyse "off-line" (Analyse exploratoire de données)

Dans cet exemple, pour simplifier l'analyse, nous n'avons pas simulé la présence d'observations manquantes ni la présence d'observations aberrantes. Pour la même raison, nous n'avons pas ajouté à l'ensemble de variables déjà mentionnées dans le précédent paragraphe des variables qui n'apportent pas d'informations utiles pour la construction des classes. Par conséquent, dans ce paragraphe nous allons seulement présenter le fonctionnement du module de filtrage et du module de classification.

Carte de contrôle T^2 de Hotelling

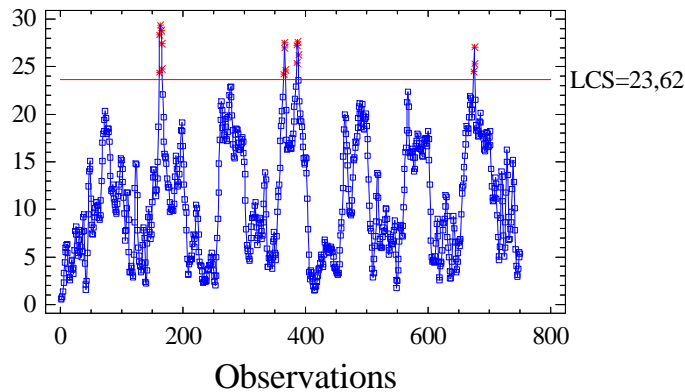


figure III.30 - Carte de contrôle T^2 pour les observations dans la base de données

Dans la figure III.30 on présente la carte de contrôle T^2 de Hotelling pour les observations simulées dans la base de données utilisateur. Malgré le grand nombre de points situés près de la limite de contrôle, seules trois des causes assignables simulées ont été détectées sur la carte. Cet exemple montre, encore une fois, combien il est difficile de détecter à l'aide de la carte de contrôle T^2 de Hotelling les causes assignables de faibles amplitudes!

Par conséquent, nous avons filtré toutes les observations contenues dans la base de données à l'aide d'un filtre passe-bas de type Butterworth d'ordre $L=2$ et ayant une fréquence de coupure $FC=0.05$ Hz. Ce filtrage nous a donné la possibilité de réduire la variabilité de variables de processus comme le montre la figure III.31 pour une portion de la courbe du débit du catalyseur dans la deuxième section.

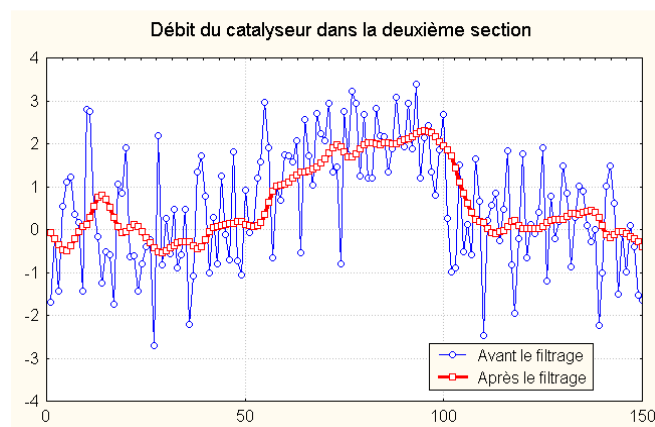


figure III.31 - débit du catalyseur dans la deuxième section

Sachant que toutes les variables sont nécessaires pour l'opération de classification, sans passer par le module de sélection des variables, nous avons effectué directement une classification non-hiérarchique. En supposant qu'on ne connaît pas le nombre réel de classes, nous avons choisi d'effectuer une classification autour des centres mobiles en fixant intentionnellement le nombre de classes à 24, c'est-à-dire trois fois plus que le nombre réel des classes dans la structure de données. Après 7 itérations nous avons obtenu la convergence de l'algorithme. Nous avons calculé, par la suite, les centres de gravités des classes ainsi formées (tableau III.12).

tableau III.12

	P	D _{S1}	D _{S2}	D _{C1}	D _{C2}	T _{max1}	T _{sortie1}	z ₁	T _{max2}	T _{sortie2}	z ₂
Classe 1	0,344	-0,107	0,104	0,346	1,951	0,190	0,265	-0,183	2,468	2,458	-2,296
Classe 2	-0,170	-0,242	2,083	-0,277	0,001	-0,292	-0,255	0,539	-0,074	-0,100	0,264
Classe 3	-0,275	-0,129	-0,501	0,157	-0,661	-2,103	-2,040	1,846	-2,360	-2,204	2,209
Classe 4	0,368	0,256	-0,223	1,916	0,424	2,304	2,203	-1,753	0,390	0,375	-0,340
Classe 5	3,757	0,024	0,220	-0,369	-0,186	1,624	1,547	-1,098	1,909	1,828	-1,507
Classe 6	0,095	0,222	1,876	-0,070	0,300	-0,123	-0,122	0,103	0,292	0,423	-0,283
Classe 7	0,084	0,250	-0,141	0,470	-0,218	0,435	0,515	-0,555	-0,238	-0,260	0,345
Classe 8	-0,027	-0,240	-0,150	0,149	-0,207	0,112	0,163	-0,022	-0,233	-0,229	0,363
Classe 9	4,025	-0,171	-0,503	-0,144	0,151	1,938	1,936	-1,419	2,120	2,145	-1,765
Classe 10	2,436	-0,394	-0,362	-0,499	0,440	0,930	0,776	-0,469	1,369	1,431	-0,929
Classe 11	0,179	0,210	-0,301	0,201	0,093	0,237	3,273	-0,247	0,157	3,185	-0,033
Classe 12	0,266	-0,188	-0,015	0,179	0,613	0,102	0,281	-0,046	0,593	0,801	-0,531
Classe 13	-0,115	2,055	-0,132	-0,026	-0,123	-0,081	-0,071	-0,056	-0,169	-0,189	0,055
Classe 14	0,060	-0,115	-0,325	0,312	-0,132	-1,282	-1,146	1,010	-1,618	-1,396	1,311
Classe 15	2,537	0,113	0,287	0,585	-0,433	1,939	2,022	-1,606	0,733	0,721	-0,274
Classe 16	-0,074	-0,365	0,448	0,012	-0,129	0,006	0,075	0,119	-0,139	-0,111	0,034
Classe 17	0,154	0,124	0,101	0,113	0,257	-1,853	-1,876	1,521	-1,762	-1,645	1,218
Classe 18	0,110	-0,222	-0,151	-0,129	0,547	-0,185	2,471	0,270	0,405	3,225	-0,269
Classe 19	0,018	0,020	0,011	-0,189	0,218	-0,089	-0,002	0,117	0,128	0,096	-0,148
Classe 20	-0,048	0,130	0,073	1,935	-0,192	2,280	2,399	-1,863	-0,314	-0,077	0,363
Classe 21	-0,173	-0,658	-0,176	-0,331	1,312	-0,225	-0,254	0,428	1,871	1,903	-1,463
Classe 22	-0,288	-0,001	0,127	-0,604	-0,088	-0,503	-0,519	0,463	-0,147	-0,086	0,207
Classe 23	0,290	0,011	-0,088	0,402	0,180	0,391	0,365	-0,510	0,258	0,360	-0,276
Classe 24	0,038	0,427	0,029	1,155	0,014	1,662	1,773	-1,067	0,153	0,166	-0,336

Sur la base de ces centres de classes, nous avons effectué une classification hiérarchique ascendante (figure III.32).

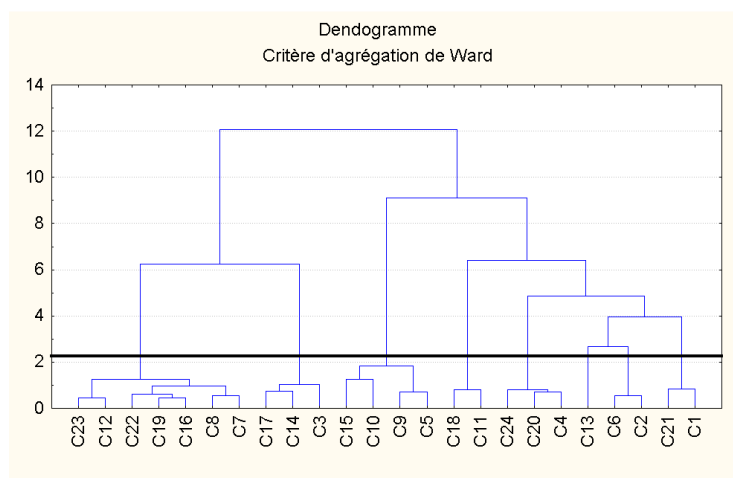


figure III.32 - dendrogramme

Pour la construction de l'arbre hiérarchique, nous avons utilisé le critère d'agrégation de Ward. La figure III.33 présente la courbe de l'indice de niveau de l'arbre qui a été utilisée pour déterminer à quel point on doit couper l'arbre. On peut remarquer facilement sur la figure III.32 que, les 24 classes se regroupent pour former finalement 8 classes (le nombre réel des classes simulées dans cet exemple).

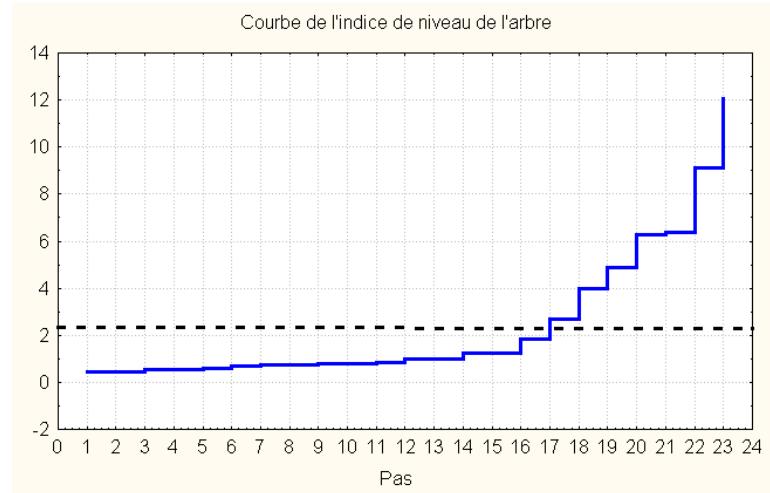


figure III.33 - courbe de l'indice de niveau

Les coordonnées des centres de gravité pour les nouvelles classes obtenues après le regroupement ont été calculées. Les résultats sont présentés dans le tableau III.13.

tableau III.13

	P	D _{S1}	D _{S2}	D _{C1}	D _{C2}	T _{max1}	T _{sortie1}	z ₁	T _{max2}	T _{sortie2}	z ₂
C₁ (C ₁ ;C ₂₁)	0,09	-0,38	-0,04	0,01	1,63	-0,02	0,01	0,12	2,17	2,18	-1,88
C₂ (C ₂ ;C ₆)	-0,04	-0,01	1,98	-0,17	0,15	-0,21	-0,19	0,32	0,11	0,16	-0,01
C₃ (C ₁₃)	-0,11	2,05	-0,13	-0,03	-0,12	-0,08	-0,07	-0,06	-0,17	-0,19	0,06
C₄ (C ₄ ;C ₂₀ ;C ₂₄)	0,12	0,27	-0,04	1,67	0,08	2,08	2,12	-1,56	0,08	0,15	-0,10
C₅ (C ₁₁ ;C ₁₈)	0,14	-0,01	-0,23	0,04	0,32	0,03	2,87	0,01	0,28	3,20	-0,15
C₆ (C ₅ ;C ₉ ;C ₁₀ ;C ₁₅)	3,19	-0,11	-0,09	-0,11	-0,01	1,61	1,57	-1,15	1,53	1,53	-1,12
C₇ (C ₃ ;C ₄ ;C ₁₇)	-0,02	-0,04	-0,24	0,19	-0,18	-1,75	-1,69	1,46	-1,91	-1,75	1,58
C₈ (C ₇ ;C ₈ ;C ₁₆ ;C ₁₉ ; C ₂₂ ;C ₁₂ ;C ₂₃)	0,04	-0,07	0,03	0,06	0,05	0,06	0,13	-0,06	0,03	0,08	0,00

Une nouvelle classification non-hiérarchique a été effectuée pour homogénéiser davantage les 8 nouvelles classes obtenues. Nous avons eu recours à une classification autour des centres de gravité présentés dans le tableau III.13. Après 4 itérations nous avons obtenu la convergence de l'algorithme et la composition finale des 8 classes. Nous avons calculé par la suite le vecteur des moyennes et les matrices de variance-covariance pour chaque classe. Ces deux paramètres ont été enregistrés dans la bibliothèque des classes.

3.5.7 Bibliothèque des classes

Afin de ne pas compliquer inutilement la présentation, nous donnons ici (tableau III.14) seulement la composition des vecteurs moyennes des classes.

tableau III.14

	P	D _{S1}	D _{S2}	D _{C1}	D _{C2}	T _{max1}	T _{sortie1}	z ₁	T _{max2}	T _{sortie2}	z ₂
C ₁	0,22	-0,24	0,04	0,20	1,77	0,10	0,16	-0,05	2,29	2,29	-2,06
C ₂	-0,13	-0,04	0,82	-0,33	0,04	-0,30	-0,28	0,33	-0,01	0,05	0,04
C ₃	0,23	0,29	-0,05	1,59	0,13	2,04	2,07	-1,51	0,17	0,21	-0,20
C ₄	0,15	0,01	-0,23	0,05	0,30	0,05	2,88	-0,02	0,27	3,18	-0,14
C ₅	3,79	-0,09	-0,17	-0,21	0,00	1,78	1,75	-1,27	1,94	1,92	-1,55
C ₆	0,05	0,04	-0,07	0,15	0,01	-1,85	-1,83	1,54	-1,89	-1,75	1,46
C ₇	0,11	-0,04	-0,08	0,22	0,02	0,21	0,28	-0,22	0,00	0,02	0,04
C ₈	-0,10	2,02	-0,13	-0,01	-0,12	-0,07	-0,06	-0,08	-0,16	-0,18	0,04

Nous avons ensuite analysé la signification physique de chaque classe en tenant compte des aspects théoriques présentés dans le paragraphe 3.5.3. Ainsi, nous avons trouvé que:

- ✓ la classe C₁ correspond à l'augmentation du débit du catalyseur dans la deuxième section du réacteur;
- ✓ la classe C₂ correspond à l'augmentation du solvant dans la deuxième section du réacteur;
- ✓ la classe C₃ correspond à l'augmentation du débit du catalyseur dans la première section du réacteur;
- ✓ la classe C₄ correspond aux dépôts de polymère sur les parois du réacteur;

- ✓ la classe C₅ correspond à l'augmentation de la pression;
- ✓ la classe C₆ correspond à l'augmentation du taux d'impuretés dans le mélange;
- ✓ la classe C₇ correspond au fonctionnement normal du processus;
- ✓ la classe C₈ correspond à l'augmentation du débit du solvant dans la première section;

3.5.8 Module d'analyse "on-line"

Après avoir constitué la bibliothèque des classes, on peut vérifier à l'aide du module d'analyse "on-line" les capacités de détection et de diagnostic de la méthodologie FNAD. Avant de commencer cette analyse, il est intéressant de remarquer que, pour les observations de la base de données (qui ont participé à la construction des axes discriminants) le pourcentage d'erreurs de classement obtenu en utilisant l'analyse discriminante est très réduit (**1,33%**), voir le tableau III.15 représentant la matrice d'erreurs de classement.

tableau III.15

	% correctes	C ₇ (norm)	C ₈ (D _{s1} ↑)	C ₂ (D _{s2} ↑)	C ₄ (Dép↑)	C ₆ (Ti↑)	C ₃ (D _{c1} ↑)	C ₁ (D _{c2} ↑)	C ₅ (P↑)
C ₇ (norm)	99,00	396	2	0	1	0	0	0	1
C ₈ (D _{s1} ↑)	98,00	1	49	0	0	0	0	0	0
C ₂ (D _{s2} ↑)	98,00	1	0	49	0	0	0	0	0
C ₄ (Dép↑)	98,00	1	0	0	49	0	0	0	0
C ₆ (Ti↑)	100,00	0	0	0	0	50	0	0	0
C ₃ (D _{c1} ↑)	98,00	1	0	0	0	0	49	0	0
C ₁ (D _{c2} ↑)	98,00	1	0	0	0	0	0	49	0
C ₅ (P↑)	98,00	1	0	0	0	0	0	0	49
Total	98,67	402	51	49	50	50	49	49	50

Après avoir analysé les données, nous avons constaté que ces erreurs apparaissent dans les zones de transition entre deux régimes de fonctionnement et sont dues à l'inertie du filtre.

Nous avons simulé également un ensemble de données de test (300 observations) contenant toutes les causes assignables précédemment décrites (tableau III.16).

tableau III.16

Causes assignables	Apparition
D _{s1} (↑) mais modérée	obs.21 - obs.40
D _{s2} (↑) mais modérée	obs.61 - obs.80
dépôts de polymère sur les parois du réacteur	obs.101 - obs.120
taux d'impuretés dans le mélange (↑)	obs.141 - obs.160
D _{c1} (↑)	obs.181 - obs.200
D _{c2} (↑)	obs.221 - obs.240
Pression: P (↑)	obs.261 - obs.280

Nous avons ensuite testé la détection de ces causes assignables à l'aide de la méthode FNAD. Les résultats obtenus sont regroupés dans le tableau III.17. Nous avons également vérifié la détection de ces causes assignables à l'aide de la carte de contrôle T^2 de Hotelling (figure III.34).

Carte de contrôle T^2 de Hotelling

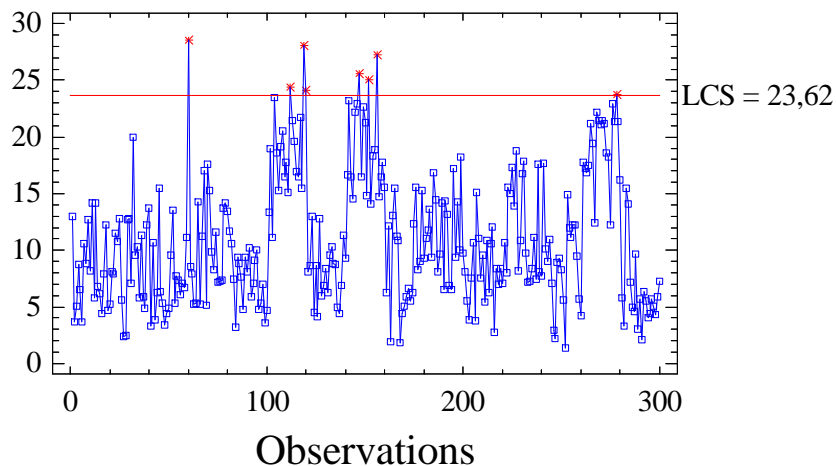


figure III.34 - la carte de contrôle Hotelling pour les observations de test

On peut remarquer que la carte de contrôle T^2 de Hotelling ne détecte pas toutes les causes assignables simulées. En effet, peu de points sur la carte dépassent la limite de contrôle et seulement trois des sept dérèglages sont détectés. Par contre, à l'aide de la méthode FNAD on détecte bien tous les dérèglages simulés. Un retard d'environ quatre observations entre le moment d'apparition du dérèglage et le moment de détection a été enregistré à cause de l'inertie du filtre. En plus, à l'aide de la méthode FNAD on est informé également de ce qui s'est réellement passé dans le processus et quelle est l'origine de la cause assignable détectée.

On peut donc conclure à travers cet exemple que la méthode FNAD est très utile pour la détection et le diagnostic des situations hors contrôle dans un processus multivarié. Le diagnostic obtenu est très précis et proche de la façon de raisonner des opérateurs car il consiste en une brève explication du phénomène qui se produit. Par la suite, il est très facile d'entreprendre les actions correctives nécessaires afin de remédier à la situation indésirable.

tableau III.17

Obs.	Simulation	Détection	Obs.	Simulation	Détection	Obs.	Simulation	Détection
1	F. Normal	F. Normal	51	F. Normal	F. Normal	101	Dép↑	F. Normal
2	F. Normal	F. Normal	52	F. Normal	F. Normal	102	Dép↑	F. Normal
3	F. Normal	F. Normal	53	F. Normal	F. Normal	103	Dép↑	F. Normal
4	F. Normal	F. Normal	54	F. Normal	F. Normal	104	Dép↑	F. Normal
5	F. Normal	F. Normal	55	F. Normal	F. Normal	105	Dép↑	F. Normal
6	F. Normal	F. Normal	56	F. Normal	F. Normal	106	Dép↑	Dép↑
7	F. Normal	F. Normal	57	F. Normal	F. Normal	107	Dép↑	Dép↑
8	F. Normal	F. Normal	58	F. Normal	F. Normal	108	Dép↑	Dép↑
9	F. Normal	F. Normal	59	F. Normal	F. Normal	109	Dép↑	Dép↑
10	F. Normal	F. Normal	60	F. Normal	F. Normal	110	Dép↑	Dép↑
11	F. Normal	F. Normal	61	D _{s2} ↑	F. Normal	111	Dép↑	Dép↑
12	F. Normal	F. Normal	62	D _{s2} ↑	F. Normal	112	Dép↑	Dép↑
13	F. Normal	F. Normal	63	D _{s2} ↑	F. Normal	113	Dép↑	Dép↑
14	F. Normal	F. Normal	64	D _{s2} ↑	D _{s2} ↑	114	Dép↑	Dép↑
15	F. Normal	F. Normal	65	D _{s2} ↑	D _{s2} ↑	115	Dép↑	Dép↑
16	F. Normal	F. Normal	66	D _{s2} ↑	D _{s2} ↑	116	Dép↑	Dép↑
17	F. Normal	F. Normal	67	D _{s2} ↑	D _{s2} ↑	117	Dép↑	Dép↑
18	F. Normal	F. Normal	68	D _{s2} ↑	D _{s2} ↑	118	Dép↑	Dép↑
19	F. Normal	F. Normal	69	D _{s2} ↑	D _{s2} ↑	119	Dép↑	Dép↑
20	F. Normal	F. Normal	70	D _{s2} ↑	D _{s2} ↑	120	Dép↑	Dép↑
21	D _{s1} ↑	F. Normal	71	D _{s2} ↑	D _{s2} ↑	121	F. Normal	Dép↑
22	D _{s1} ↑	F. Normal	72	D _{s2} ↑	D _{s2} ↑	122	F. Normal	Dép↑
23	D _{s1} ↑	F. Normal	73	D _{s2} ↑	D _{s2} ↑	123	F. Normal	Dép↑
24	D _{s1} ↑	F. Normal	74	D _{s2} ↑	D _{s2} ↑	124	F. Normal	Dép↑
25	D _{s1} ↑	D _{s1} ↑	75	D _{s2} ↑	D _{s2} ↑	125	F. Normal	F. Normal
26	D _{s1} ↑	D _{s1} ↑	76	D _{s2} ↑	D _{s2} ↑	126	F. Normal	F. Normal
27	D _{s1} ↑	D _{s1} ↑	77	D _{s2} ↑	D _{s2} ↑	127	F. Normal	F. Normal
28	D _{s1} ↑	D _{s1} ↑	78	D _{s2} ↑	D _{s2} ↑	128	F. Normal	F. Normal
29	D _{s1} ↑	D _{s1} ↑	79	D _{s2} ↑	D _{s2} ↑	129	F. Normal	F. Normal
30	D _{s1} ↑	D _{s1} ↑	80	D _{s2} ↑	D _{s2} ↑	130	F. Normal	F. Normal
31	D _{s1} ↑	D _{s1} ↑	81	F. Normal	D _{s2} ↑	131	F. Normal	F. Normal
32	D _{s1} ↑	D _{s1} ↑	82	F. Normal	D _{s2} ↑	132	F. Normal	F. Normal
33	D _{s1} ↑	D _{s1} ↑	83	F. Normal	D _{s2} ↑	133	F. Normal	F. Normal
34	D _{s1} ↑	D _{s1} ↑	84	F. Normal	D _{s2} ↑	134	F. Normal	F. Normal
35	D _{s1} ↑	D _{s1} ↑	85	F. Normal	F. Normal	135	F. Normal	F. Normal
36	D _{s1} ↑	D _{s1} ↑	86	F. Normal	F. Normal	136	F. Normal	F. Normal
37	D _{s1} ↑	D _{s1} ↑	87	F. Normal	F. Normal	137	F. Normal	F. Normal
38	D _{s1} ↑	D _{s1} ↑	88	F. Normal	F. Normal	138	F. Normal	F. Normal
39	D _{s1} ↑	D _{s1} ↑	89	F. Normal	F. Normal	139	F. Normal	F. Normal
40	D _{s1} ↑	D _{s1} ↑	90	F. Normal	F. Normal	140	F. Normal	F. Normal
41	F. Normal	D _{s1} ↑	91	F. Normal	F. Normal	141	Ti↑	F. Normal
42	F. Normal	D _{s1} ↑	92	F. Normal	F. Normal	142	Ti↑	F. Normal
43	F. Normal	D _{s1} ↑	93	F. Normal	F. Normal	143	Ti↑	F. Normal
44	F. Normal	F. Normal	94	F. Normal	F. Normal	144	Ti↑	F. Normal
45	F. Normal	F. Normal	95	F. Normal	F. Normal	145	Ti↑	F. Normal
46	F. Normal	F. Normal	96	F. Normal	F. Normal	146	Ti↑	Ti↑
47	F. Normal	F. Normal	97	F. Normal	F. Normal	147	Ti↑	Ti↑
48	F. Normal	F. Normal	98	F. Normal	F. Normal	148	Ti↑	Ti↑
49	F. Normal	F. Normal	99	F. Normal	F. Normal	149	Ti↑	Ti↑
50	F. Normal	F. Normal	100	F. Normal	F. Normal	150	Ti↑	Ti↑

tableau III.17 (continuation)

Obs.	Simulation	Détection	Obs.	Simulation	Détection	Obs.	Simulation	Détection
151	Ti↑	Ti↑	201	F. Normal	D _{c1} ↑	251	F. Normal	F. Normal
152	Ti↑	Ti↑	202	F. Normal	D _{c1} ↑	252	F. Normal	F. Normal
153	Ti↑	Ti↑	203	F. Normal	D _{c1} ↑	253	F. Normal	F. Normal
154	Ti↑	Ti↑	204	F. Normal	D _{c1} ↑	254	F. Normal	F. Normal
155	Ti↑	Ti↑	205	F. Normal	F. Normal	255	F. Normal	F. Normal
156	Ti↑	Ti↑	206	F. Normal	F. Normal	256	F. Normal	F. Normal
157	Ti↑	Ti↑	207	F. Normal	F. Normal	257	F. Normal	F. Normal
158	Ti↑	Ti↑	208	F. Normal	F. Normal	258	F. Normal	F. Normal
159	Ti↑	Ti↑	209	F. Normal	F. Normal	259	F. Normal	F. Normal
160	Ti↑	Ti↑	210	F. Normal	F. Normal	260	F. Normal	F. Normal
161	F. Normal	Ti↑	211	F. Normal	F. Normal	261	P↑	F. Normal
162	F. Normal	Ti↑	212	F. Normal	F. Normal	262	P↑	F. Normal
163	F. Normal	Ti↑	213	F. Normal	F. Normal	263	P↑	F. Normal
164	F. Normal	Ti↑	214	F. Normal	F. Normal	264	P↑	F. Normal
165	F. Normal	Ti↑	215	F. Normal	F. Normal	265	P↑	F. Normal
166	F. Normal	F. Normal	216	F. Normal	F. Normal	266	P↑	P↑
167	F. Normal	F. Normal	217	F. Normal	F. Normal	267	P↑	P↑
168	F. Normal	F. Normal	218	F. Normal	F. Normal	268	P↑	P↑
169	F. Normal	F. Normal	219	F. Normal	F. Normal	269	P↑	P↑
170	F. Normal	F. Normal	220	F. Normal	F. Normal	270	P↑	P↑
171	F. Normal	F. Normal	221	D _{c2} ↑	F. Normal	271	P↑	P↑
172	F. Normal	F. Normal	222	D _{c2} ↑	F. Normal	272	P↑	P↑
173	F. Normal	F. Normal	223	D _{c2} ↑	F. Normal	273	P↑	P↑
174	F. Normal	F. Normal	224	D _{c2} ↑	F. Normal	274	P↑	P↑
175	F. Normal	F. Normal	225	D _{c2} ↑	F. Normal	275	P↑	P↑
176	F. Normal	F. Normal	226	D _{c2} ↑	D _{c2} ↑	276	P↑	P↑
177	F. Normal	F. Normal	227	D _{c2} ↑	D _{c2} ↑	277	P↑	P↑
178	F. Normal	F. Normal	228	D _{c2} ↑	D _{c2} ↑	278	P↑	P↑
179	F. Normal	F. Normal	229	D _{c2} ↑	D _{c2} ↑	279	P↑	P↑
180	F. Normal	F. Normal	230	D _{c2} ↑	D _{c2} ↑	280	P↑	P↑
181	D _{c1} ↑	F. Normal	231	D _{c2} ↑	D _{c2} ↑	281	F. Normal	P↑
182	D _{c1} ↑	F. Normal	232	D _{c2} ↑	D _{c2} ↑	282	F. Normal	P↑
183	D _{c1} ↑	F. Normal	233	D _{c2} ↑	D _{c2} ↑	283	F. Normal	P↑
184	D _{c1} ↑	F. Normal	234	D _{c2} ↑	D _{c2} ↑	284	F. Normal	P↑
185	D _{c1} ↑	F. Normal	235	D _{c2} ↑	D _{c2} ↑	285	F. Normal	P↑
186	D _{c1} ↑	D _{c1} ↑	236	D _{c2} ↑	D _{c2} ↑	286	F. Normal	F. Normal
187	D _{c1} ↑	D _{c1} ↑	237	D _{c2} ↑	D _{c2} ↑	287	F. Normal	F. Normal
188	D _{c1} ↑	D _{c1} ↑	238	D _{c2} ↑	D _{c2} ↑	288	F. Normal	F. Normal
189	D _{c1} ↑	D _{c1} ↑	239	D _{c2} ↑	D _{c2} ↑	289	F. Normal	F. Normal
190	D _{c1} ↑	D _{c1} ↑	240	D _{c2} ↑	D _{c2} ↑	290	F. Normal	F. Normal
191	D _{c1} ↑	D _{c1} ↑	241	F. Normal	D _{c2} ↑	291	F. Normal	F. Normal
192	D _{c1} ↑	D _{c1} ↑	242	F. Normal	D _{c2} ↑	292	F. Normal	F. Normal
193	D _{c1} ↑	D _{c1} ↑	243	F. Normal	D _{c2} ↑	293	F. Normal	F. Normal
194	D _{c1} ↑	D _{c1} ↑	244	F. Normal	F. Normal	294	F. Normal	F. Normal
195	D _{c1} ↑	D _{c1} ↑	245	F. Normal	F. Normal	295	F. Normal	F. Normal
196	D _{c1} ↑	D _{c1} ↑	246	F. Normal	F. Normal	296	F. Normal	F. Normal
197	D _{c1} ↑	D _{c1} ↑	247	F. Normal	F. Normal	297	F. Normal	F. Normal
198	D _{c1} ↑	D _{c1} ↑	248	F. Normal	F. Normal	298	F. Normal	F. Normal
199	D _{c1} ↑	D _{c1} ↑	249	F. Normal	F. Normal	299	F. Normal	F. Normal
200	D _{c1} ↑	D _{c1} ↑	250	F. Normal	F. Normal	300	F. Normal	F. Normal

3.6 Conclusion

Nous avons présenté au début du chapitre quelques notions théoriques concernant l'aspect descriptif et l'aspect décisionnel de la méthode d'analyse discriminante. Un intérêt particulier a été porté aux caractéristiques qualitatives des règles de classement. Ensuite, nous avons passé en revue différents types de filtres numériques et nous avons précisé les équations récurrentes les caractérisant.

Nous avons proposé une nouvelle méthode de contrôle en temps réel – FNAD - basée sur l'utilisation conjointe de deux outils: le filtrage numérique et l'analyse discriminante. Une fois mise en œuvre, cette méthode de contrôle de processus est très simple à exploiter et les résultats obtenus sont très facilement interprétables. Nous avons intégré dans notre démarche toute l'information disponible concernant le processus (l'avis des experts et l'analyse historiques des données disponibles dans une base de données utilisateur). L'objectif de la méthode FNAD est double: détecter les différents dérèglages (ou causes assignables) susceptibles d'influencer le processus et identifier le cas échéant la variable ou les variables à l'origine d'un fonctionnement anormal du processus. Nous avons effectué, d'abord, une analyse détaillée de deux causes assignables, souvent rencontrées en pratique: le saut en échelon et la dérive en rampe. Ces deux types de causes assignables ont été modélisées et analysées dans le temps et en fréquence. Ensuite, nous avons décrit les différents modules dans la composition du schéma de la méthode FNAD. Un intérêt particulier a été accordé au module de filtrage numérique et au module d'analyse discriminante. Les caractéristiques et les performances des différents filtres numériques ont été analysées afin de déterminer le filtre le plus adéquat pour les objectifs que nous nous sommes fixés. L'efficacité de l'analyse discriminante en tant qu'outil d'aide à la prise de décision a été analysée dans le cas des variables indépendantes et corrélées.

Finalement, un exemple pratique d'utilisation de la méthode a été donné pour le diagnostic d'un processus de fabrication du polyéthylène de faible densité. A travers cet exemple nous avons testé le fonctionnement de la plupart des modules constituant le schéma fonctionnel de la méthode.

Chapitre 4

- 4.1 Introduction
- 4.2 Carte de contrôle par filtrage
- 4.3 Carte de contrôle spectrale
- 4.4 Conclusion

4.1 Introduction

Dans ce chapitre nous aborderons quelques perspectives ouvertes dans le domaine de la MSP par l'introduction des techniques et des outils qui ne sont pas forcément propres à la MSP mais qui sont fréquemment utilisés dans d'autres domaines comme, par exemple, le traitement numérique de signal, la spectrographie etc...

Le premier aspect analysé est lié à l'utilisation des filtres numériques pour la détection des dérèglages de faible amplitude de la moyenne dans une série de données. Une nouvelle carte de contrôle dont la construction repose sur des équations récurrentes définissant les filtres numériques RII (Réponse Impulsionnelle Infinie) sera proposée. Les performances de cette nouvelle carte de contrôle, nommée Carte de Contrôle par Filtrage Numérique (CCFN), seront analysées et une comparaison sera effectuée avec la carte de contrôle EWMA. Les difficultés liées au calcul théorique de la Période Opérationnelle Moyenne de la CCFN seront exposées.

Le deuxième aspect traité dans ce chapitre est également lié à la détection des changements de la moyenne dans une série de données, en utilisant comme outil l'analyse spectrale. Nous allons analyser la dualité temps-fréquence et nous proposerons une nouvelle carte de contrôle, nommée Carte de Contrôle Spectrale (CCS), dont le principe repose sur l'analyse du spectre. Les performances de la carte de contrôle CCS dans la détection des dérèglages seront analysées. La POM (Période Opérationnelle Moyenne) de la nouvelle carte de contrôle sera comparée à celle de la carte de contrôle EWMA.

4.2 Carte de Contrôle par Filtrage Numérique (CCFN)

4.2.1 Introduction

Nous avons présenté dans le premier chapitre la contribution majeure des cartes de contrôle à l'amélioration de la qualité des processus industriels ou d'autres natures. Le principal inconvénient de la carte de contrôle de type Shewhart est qu'elle a des performances modestes dans la détection des dérèglages de faibles amplitudes de la moyenne. Cela est dû au fait qu'elle ne tient compte que de l'observation courante. Pour améliorer la détection de ce type de dérèglages, des cartes de contrôle plus performantes, tenant compte d'une série des valeurs, ont été proposées. On rappelle ici, à titre d'exemple, les cartes de contrôle EWMA de Roberts [103] et la carte de contrôle CUSUM de Page [90]. Nous allons rappeler brièvement quelques notions théoriques concernant la carte de contrôle EWMA qui vont nous servir pour une comparaison ultérieure avec la carte de contrôle que l'on propose, la carte CCFN.

La carte de contrôle EWMA est basée sur la statistique:

$$y_k = r x_k + (1-r)y_{k-1}, \quad 0 < r \leq 1 \text{ et } k = 0, 1, \dots \quad (4.1)$$

On prend souvent comme valeur de démarrage y_0 la valeur cible du procédé ($y_0 = \mu_0$).. Les observations séquentielles $y = \{y_k \mid k = 0, 1, \dots\}$ enregistrées sur la carte peuvent être des valeurs individuelles ou des moyennes d'échantillons. Par une transformation très simple de l'équation (4.1), on peut exprimer la valeur de y_k en fonction de la valeur courante x_k et des valeurs précédentes $\{x_{k-j} \mid j=1, 2, \dots, k-1\}$ sous la forme:

$$y_k = r \sum_{j=0}^{k-1} (1-r)^j x_{k-j} + (1-r)^k y_0 \quad (4.2)$$

et on remarque que les poids associés aux observations antérieures diminuent exponentiellement comme dans une série géométrique.

Si les valeurs $x = \{x_k \mid k = 1, 2, \dots\}$ sont indépendantes et identiquement distribuées (iid), alors la variance de y_k peut s'exprimer en fonction de la variance de x sous la forme:

$$\sigma_{y_k}^2 = [1 - (1-r)^{2k}] \frac{r}{2-r} \sigma_x^2 \quad (4.3)$$

Sauf pour le cas où la valeur de la constante r est très faible, la variance de y_k converge rapidement vers $\sigma_y^2 = \frac{r}{2-r} \sigma_x^2$ et donc, les limites de contrôle de la carte EWMA pourront être calculées à l'aide des formules ci-dessous:

$$\begin{aligned} \text{LSC} &= \mu_0 + h_{\text{EWMA}} \sigma_x \sqrt{\frac{r}{2-r}} \\ \text{LIC} &= \mu_0 - h_{\text{EWMA}} \sigma_x \sqrt{\frac{r}{2-r}} \end{aligned} \quad (4.4)$$

où le paramètre h_{EWMA} est choisi afin de satisfaire une certaine POM_0 (Période Opérationnelle Moyenne sous contrôle).

4.2.2 Principe de construction de la carte CCFN

Le principe de construction de toutes les cartes de contrôle est basé sur un test d'hypothèse. On doit décider entre deux hypothèses:

- ✓ H_0 : "la caractéristique suivie n'a pas changé"
- ✓ H_1 : "la caractéristique suivie a changé"

Pour cela une certaine statistique de test est définie et un critère de décision est établi. Deux types d'erreurs sont possibles:

- ✓ décider que l'hypothèse H_0 est fautive alors qu'elle est vraie (risque α ou risque de première espèce)
- ✓ décider que l'hypothèse H_0 est vraie alors qu'elle est fautive (risque β ou risque de deuxième espèce).

Pour les cartes de contrôle, le risque α est le risque d'avoir des fausses alertes et évidemment on le veut le plus petit possible (voir figure IV.1.a). Ce risque est lié à la POM_0 (Période Opérationnelle Moyenne sous Contrôle) de la carte par l'équation $POM_0 = 1/\alpha$.

Le risque β représente le risque de ne pas détecter sur la carte le dérèglement et on le veut aussi le plus petit possible (voir figure IV.1.b). Il est lié à la POM_1 (Période Opérationnelle Moyenne hors-contrôle) par l'équation $POM_1 = 1/(1 - \beta)$.

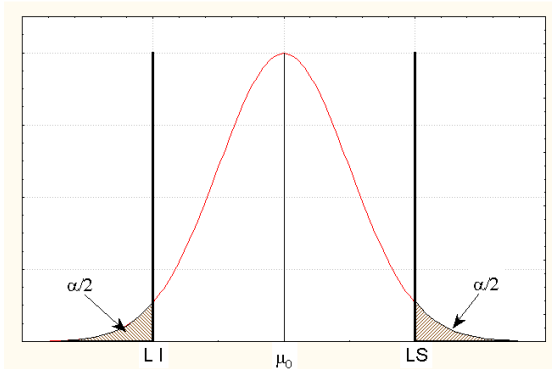


figure IV.1.a - risque α

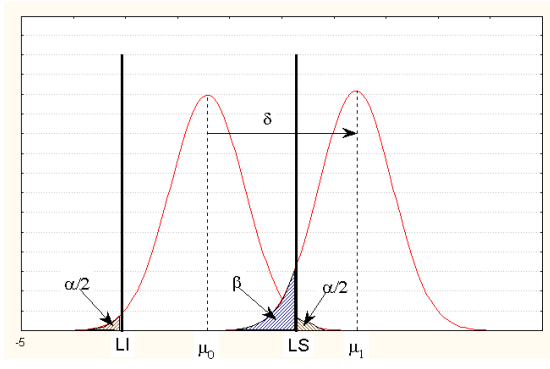


figure IV.1.b - risque β

Les limites de contrôle de la carte sont choisies afin de ne pas dépasser un risque α imposé. En effet, le risque α accepté est fixé en priorité, car si l'opérateur qui surveille le processus fait un réglage suite à une fausse alerte sur la carte de contrôle, il va augmenter artificiellement la variabilité du processus et cela vient en contradiction avec l'objectif énoncé d'une carte de contrôle. Pour un dérèglement donné et un risque α fixé, le risque β de la carte de contrôle est facilement calculable.

tableau IV.1

$\delta \backslash n$	1	2	3	4	5	10	15	20
0.5	9,94E-01	9,89E-01	9,84E-01	9,77E-01	9,70E-01	9,22E-01	8,56E-01	7,78E-01
0.75	9,88E-01	9,74E-01	9,56E-01	9,33E-01	9,07E-01	7,35E-01	5,38E-01	3,62E-01
1	9,77E-01	9,44E-01	8,98E-01	8,41E-01	7,78E-01	4,36E-01	1,91E-01	7,05E-02
1.25	9,60E-01	8,91E-01	7,98E-01	6,91E-01	5,81E-01	1,70E-01	3,28E-02	4,80E-03
1.5	9,33E-01	8,10E-01	6,56E-01	5,00E-01	3,62E-01	4,06E-02	2,48E-03	1,04E-04
1.75	8,94E-01	7,00E-01	4,88E-01	3,09E-01	1,81E-01	5,64E-03	7,91E-05	6,96E-07
2	8,41E-01	5,68E-01	3,21E-01	1,59E-01	7,05E-02	4,43E-04	1,04E-06	1,39E-09
2.25	7,73E-01	4,28E-01	1,85E-01	6,68E-02	2,11E-02	1,93E-05	5,51E-09	8,19E-13
2.5	6,91E-01	2,96E-01	9,17E-02	2,28E-02	4,80E-03	4,65E-07	1,17E-11	1,42E-16
2.75	5,99E-01	1,87E-01	3,89E-02	6,21E-03	8,19E-04	6,12E-09	9,99E-15	7,13E-21
3	5,00E-01	1,07E-01	1,40E-02	1,35E-03	1,04E-04	4,38E-11	3,38E-18	1,04E-25

Dans le tableau IV.1 il est présenté le risque β pour une carte de contrôle de type Shewhart ayant les limites de contrôle fixées à ± 3 écarts-type ($\alpha = 0,27\%$), en fonction de la taille de l'échantillon et de l'amplitude du dérèglement (en multiples d'écart-type). On donne dans la figure IV.2 une représentation graphique du risque β dont les valeurs sont exposées dans le tableau IV.1.

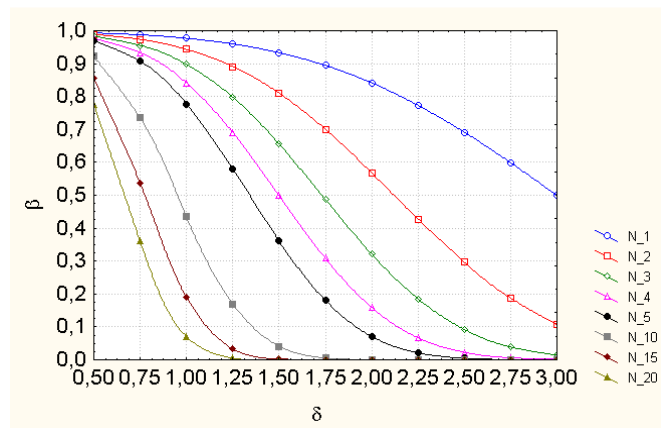


figure IV.2 – risque β en fonction de l'amplitude du dérèglement et de la taille d'échantillon pour une carte de contrôle Shewhart ayant les limites de contrôle fixées à ± 3 écarts-type ($\alpha = 0,27\%$)

On remarque, par exemple, que dans le cas des observation individuelles (taille d'échantillon $n = 1$), le risque β de la carte de contrôle est très élevé surtout quand l'amplitude du dérèglement est faible.

Pour remédier à cet inconvénient deux solutions sont possibles. La première consiste à réduire l'intervalle entre les deux limites de contrôle (figure IV.1.b) ce qui se traduit par la réduction du risque β mais aussi par l'augmentation du risque α . Or, nous avons vu que le risque α est le risque d'avoir des fausses alertes sur la carte de contrôle, ce qui n'est pas souhaitable! Donc cette solution n'est pas très convenable.

Une deuxième solution consiste à augmenter la taille de l'échantillon et suivre la carte des moyennes. L'augmentation de la taille de l'échantillon se traduit par la réduction de l'écart-type de la distribution des moyennes selon la formule:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \tag{4.5}$$

où $\sigma_{\bar{x}}$ est l'écart-type de la distribution des moyennes et σ_x est l'écart-type de la caractéristique analysée. Ainsi pour le même risque α et le même dérèglement δ le risque β

est réduit (figure IV.1.b). Cette deuxième solution est clairement meilleure que la première mais il ne faut pas oublier que même si, finalement, on détecte le dérèglement plus vite en utilisant la carte de contrôle des moyennes \bar{X} , le nombre des observations individuelles nécessaires afin de détecter le dérèglement reste élevé.

Pour conclure, on peut remarquer dans le cas des cartes de contrôle, l'importance de la réduction de la variabilité de la caractéristique étudiée pour la réduction du risque β et donc pour l'augmentation de la POM_1 ce qui veut dire une vitesse de réaction plus élevée à la présence des causes assignables.

En partant de ce constat, on peut se demander s'il n'y a pas un autre moyen plus efficace de réduction de la variabilité d'une caractéristique que l'utilisation de la moyenne des échantillons. Dans le chapitre 3, nous avons étudié la composition spectrale du saut en échelon et de la dérive en rampe, deux types de causes assignables fréquemment rencontrées dans la pratique. Nous avons montré que leur présence se manifeste au préalable à des basses fréquences. Par contre, les causes communes qui peuvent être considérées comme un bruit blanc ont des composantes spectrales uniformément distribuées dans toute la bande de fréquences.

Si on veut mettre plus clairement en évidence la cause assignable, il faudra utiliser des filtres numériques passe-bas. Le filtre passe-bas laisse passer uniquement les basse-fréquences, donc surtout les composantes de la cause assignable et une faible partie de la composition spectrale du bruit. Par contre, les composantes spectrales de hautes fréquences qui sont presque en totalité les composantes du bruit seront éliminées. De cette manière la dispersion de la caractéristique étudiée est réduite. On peut dire que le filtre numérique agit comme un véritable estimateur de la moyenne de la caractéristique suivie.

Nous avons défini au chapitre 3 le filtrage numérique de signal comme une combinaison linéaire des échantillons de l'entrée et de la sortie d'un système échantillonné et nous avons donné les fonctions de transfert pour les différents filtres numériques. Nous reprenons ici, en utilisant la transformée en z , la formule de la fonction de transfert ARMA (AutoRegressive Moving Average) d'un filtre numérique qui est un rapport des deux polynômes en z :

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{i=0}^n a_i z^{-i}}{1 - \sum_{j=1}^m b_j z^{-j}} \quad (4.6)$$

Si on revient dans le domaine temporel, en utilisant les propriétés de la transformée en z , on obtient l'équation de la sortie du filtre à l'instant k en fonction de l'état de l'entrée à l'instant k et de l'état d'un certain nombre d'entrées et de sorties à des instants antérieurs:

$$b_0 y_k = \sum_{i=0}^n a_i x_{k-i} + \sum_{j=1}^m b_j y_{k-j} \quad (4.7)$$

D'une façon générale, pour simplifier, on prend $b_0 = 1$. Une représentation schématique du filtre ARMA est donnée dans la figure IV.3.

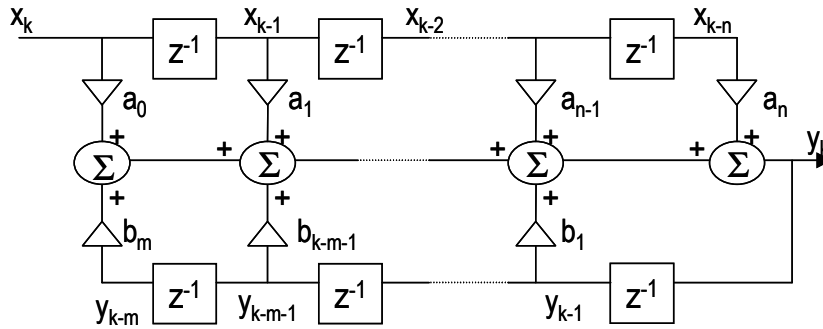


figure IV.3 – Schéma du filtre ARMA

On rappelle, voir le chapitre 3, que les coefficients a_i et b_j dans l'équation (4.6) sont calculés pour assurer une certaine réponse en fréquence du filtre numérique ou l'équivalence avec un modèle de filtre analogique. En ce qui concerne la réalisation pratique des filtres, deux contraintes majeures doivent être respectées: la *causalité* ($h_n = 0$ pour $n < 0$, où $[h_n]$ définit la réponse impulsionnelle du filtre) et la *stabilité* (les zéros du dénominateur dans l'équation (4.6) doivent être situés à l'intérieur du cercle unité dans le plan z).

En tenant compte du fait que, pour la fréquence nulle ($f=0$), c'est-à-dire pour $z = 1$, le gain du filtre doit être unitaire, l'équation (4.6) nous donne dans le cas général:

$$\frac{\sum_{i=0}^n a_i}{1 - \sum_{j=1}^m b_j} = 1 \Leftrightarrow \sum_{i=0}^n a_i = 1 - \sum_{j=1}^m b_j \quad (4.8)$$

A partir de l'équation (4.7) en considérant $b_0=1$, on calcule l'espérance mathématique de y en fonction de l'espérance mathématique de x :

$$\begin{aligned} E[y_k] &= E\left[\sum_{i=0}^n a_i x_{k-i} + \sum_{j=1}^m b_j y_{k-j}\right] = E\left[\sum_{i=0}^n a_i x_{k-i}\right] + E\left[\sum_{j=1}^m b_j y_{k-j}\right] = \sum_{i=0}^n a_i E[x_{k-i}] + \\ &+ \sum_{j=1}^m b_j E[y_{k-j}] \end{aligned}$$

d'où

$$\mu_y = \mu_x \sum_{i=0}^n a_i + \mu_y \sum_{j=1}^m b_j \Rightarrow \mu_y = \frac{\sum_{i=0}^n a_i}{1 - \sum_{j=1}^m b_j} \mu_x = \mu_x \quad (\text{selon l'équation 4.8}). \quad (4.9)$$

Donc la moyenne du signal à la sortie du filtre μ_y est égale à la moyenne du signal à l'entrée du filtre μ_x .

Pour calculer la variance du signal à la sortie du filtre σ_y^2 il va falloir transformer l'équation récurrente (4.7) pour exprimer y en fonction de x seulement.

Pour simplifier, considérons que $n = m = L$ (l'ordre du filtre). Après des transformations algébriques élémentaires (annexe A.5), l'équation (4.7) peut s'écrire sous la forme suivante:

$$y_k = \sum_{i=0}^k c_i x_{k-i}, \quad \text{où } c_i = \begin{cases} a_i + \sum_{j=1}^i b_j c_{i-j} & , i \leq L \\ \sum_{j=1}^L b_j c_{i-j} & , i > L \end{cases} \quad (4.10)$$

et donc:

$$\sigma_y^2 = \text{Var}[y_k] = \text{Var}\left[\sum_{i=1}^k c_i x_{k-i}\right] = \sum_{i=1}^k \text{Var}[c_i x_{k-i}] = \sum_{i=1}^k c_i^2 \text{Var}[x_{k-i}] = \sigma_x^2 \left(\sum_{i=1}^k c_i^2\right) \quad (4.11)$$

La somme $\sum_{i=1}^k c_i^2$ converge toujours vers une valeur constante qui est déterminée selon le type de filtre utilisé et de sa fréquence de coupure.

Dans le cas des filtres passe-bas, pour le même ordre du filtre L la convergence de la somme précédente est d'autant plus rapide que la fréquence de coupure du filtre augmente. Nous présentons dans le tableau IV.2, la constante vers laquelle converge cette somme pour un filtre passe-bas d'ordre $L=2$ de type Butterworth en fonction de la fréquence de coupure du filtre (entre crochets on a mentionné le nombre approximatif de valeurs jusqu'à la convergence de la série).

Tableau IV.2

F_c	0.01	0.03	0.04	0.06	0.08	0.1	0.2
$C = \sum_{i=1}^k c_i^2$	0.022 [54]	0.066 [18]	0.088 [15]	0.131 [10]	0.173 [10]	0.214 [7]	0.407 [7]

Nous avons constaté qu'il y a pour ce type de filtre une dépendance linéaire entre la fréquence de coupure et la constante C (figure IV.4), et nous avons déduit l'équation suivante: $C = 0.007 + 2.02 F_c$, qui nous permettra de déterminer la constante C en fonction de la fréquence de coupure du filtre pour tout filtre passe-bas d'ordre $L=2$ de type Butterworth. Une telle équation pourra être déterminée pour tout autre type de filtre numérique.

Par conséquent, selon les équations (4.9) et (4.11) nous proposons une nouvelle carte de contrôle basée sur le filtrage numérique que l'on convient d'appeler CCFN (Carte de Contrôle par Filtrage Numérique). Sur cette carte de contrôle, on va représenter la caractéristique y_k définie dans l'équation (4.7) en fonction de l'instant k . La carte de contrôle CCFN devrait avoir les limites de contrôle suivantes:

$$LC_y = \mu_y \pm h_{CCFN} \sigma_y$$

et tenant compte de l'équation (4.9) et de l'équation (4.10) cela revient à fixer les limites de contrôle à:

$$\begin{aligned} \text{LSC} &= \mu_x + h_{\text{CCFN}} \sigma_x \sqrt{C} \\ \text{LIC} &= \mu_x - h_{\text{CCFN}} \sigma_x \sqrt{C} \end{aligned} \quad (4.12)$$

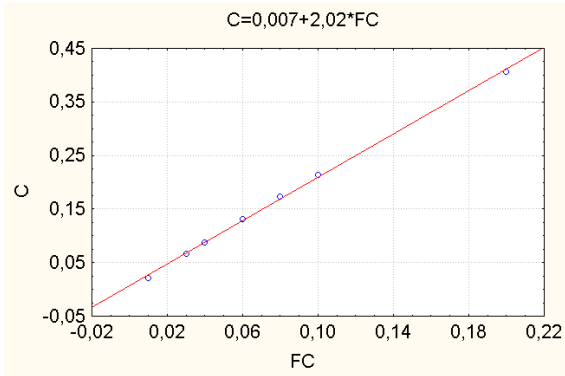


figure IV.4 – valeurs de convergence pour la somme $\sum_{i=1}^k c_i^2$ dans l'équation (4.11) en fonction de la fréquence de coupure du filtre passe-bas d'ordre $L = 2$ de type Butterworth

Il ne reste qu'à déterminer le paramètre h_{CCFN} afin d'assurer une certaine POM_0 et de calculer ensuite les performances de la carte dans la détection des différents dérèglages c'est-à-dire la POM_1 .

4.2.3 Carte multidimensionnelle de contrôle par filtrage numérique CMCFN

Le principe de construction de la carte de contrôle par filtrage présenté dans le précédent paragraphe peut être facilement transposé pour le cas multivarié. Dans ce cas, au lieu d'avoir une seule variable x_k à l'entrée du filtre et une seule variable y_k à la sortie du filtre on aura un vecteur X_k de dimension $(p \times 1)$ à l'entrée et un vecteur Y_k de dimension $(p \times 1)$ à la sortie (figure IV.5). Dans ce qui suit, on convient de noter les vecteurs avec des lettres en majuscules et les scalaires avec des lettres en minuscules.



Figure IV.5.a - Filtre SISO (Single Input Single Output)



figure IV.5.b - Filtre MIMO (Multiple Input Multiple Output)

L'équation (4.10) définissant la sortie du filtre y en fonction de l'entrée x dans le cas univarié, sera dans le cas multivarié de la forme suivante:

$$Y_k = \sum_{i=0}^k C_i X_{k-i} \quad (4.13)$$

où $Y_k = (y_{1k}, y_{2k}, \dots, y_{pk})^T$ est le vecteur de la sortie du filtre;
 $X_k = (x_{1k}, x_{2k}, \dots, x_{pk})^T$ est le vecteur de l'entrée du filtre;

$$C_i = \begin{pmatrix} c_{i1} & 0 & \cdots & 0 \\ 0 & c_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_{ip} \end{pmatrix} = c_i I_{(p \times p)} \text{ est la matrice de pondération avec:}$$

$$c_{ir} = \begin{cases} a_{ir} + \sum_{j=1}^i b_{jr} c_{i-j,r} & , i \leq L \\ \sum_{j=1}^L b_{jr} c_{i-j,r} & , i > L \end{cases}$$

où $r = 1, \dots, p$ et $I_{(p \times p)}$ est la matrice identité d'ordre $p \times p$.

La matrice de variance-covariance de la sortie Y dans le cas multivarié est de la forme suivante:

$$\Sigma_{Y_k} = \sum_{i=0}^k \text{var}[C_i X_{k-i}] = \sum_{i=0}^k [C_i X_{k-i} X_{k-i}^T C_i^T] \quad (4.14)$$

Quand il n'y a pas de raison particulière pour pondérer différemment chaque entrée du filtre, on peut considérer tous les coefficients c_{ik} de la matrice C_i égaux. Il en résulte que $C_i = C_i^T = c_i \times I_{(p \times p)}$ et l'équation (4.14) peut s'écrire de la manière suivante:

$$\Sigma_{Y_k} = \sum_{i=0}^k c_i^2 (X_{k-i} X_{k-i}^T) = \sum_{i=0}^k c_i^2 \Sigma \quad (4.15)$$

où Σ est la matrice variance-covariance de X .

La carte de contrôle multivariée va signaler à l'instant k une situation hors-contrôle si la statistique:

$$T_k^2 = Y_k^T \Sigma_{Y_k}^{-1} Y_k > h_{CMCFN} \quad (4.16)$$

où la constante positive h_{CMCFN} est choisie pour respecter une certaine période POM_0 et Σ_{Y_k} est la matrice variance-covariance de Y_k .

Nous avons calculé à partir des simulations le coefficient h_{CMCFN} intervenant dans l'équation (4.16) afin de respecter une période opérationnelle (POM_0) comprise entre $POM_0=100$ et $POM_0=1000$. Le calcul du coefficient h_{CMCFN} a été fait dans le cas des filtres passe-bas de type Butterworth d'ordre $L=2$ et ayant des fréquences de coupure comprises entre 0.01 Hz et 0.2 Hz. Les résultats obtenus, pour 300 simulations de 5000 observations chacune, sont présentés dans le tableau IV.3.

tableau IV.3 - le coefficient h_{CMCFN} en fonction de p , POM_0 et FC

N° Variables	FC	0.01	0.03	0.04	0.06	0.08	0.1	0.2
	POM_0							
p=2	100	0.20	0.61	0.81	1.20	1.59	1.96	3.74
	200	0.23	0.70	0.92	1.38	1.83	2.27	4.30
	500	0.26	0.82	1.10	1.61	2.16	2.67	5.10
	1000	0.28	0.90	1.20	1.79	2.40	2.97	5.67
p=3	100	0.25	0.75	0.99	1.48	1.95	2.43	4.64
	200	0.28	0.84	1.12	1.69	2.22	2.74	5.21
	500	0.31	0.96	1.30	1.94	2.56	3.15	6.05
	1000	0.33	1.04	1.44	2.12	2.82	3.48	6.64
p=4	100	0.28	0.87	1.16	1.73	2.30	2.90	5.40
	200	0.31	0.98	1.29	1.95	2.58	3.17	6.04
	500	0.35	1.12	1.46	2.22	2.93	3.62	6.88
	1000	0.37	1.21	1.58	2.42	3.20	3.93	7.59
p=5	100	0.33	0.99	1.31	1.97	2.61	3.22	6.12
	200	0.36	1.09	1.44	2.18	2.89	3.57	6.80
	500	0.40	1.24	1.64	2.48	3.26	4.03	7.72
	1000	0.41	1.34	1.77	2.69	3.56	4.38	8.41
p=10	100	0.49	1.52	2.03	3.03	4.00	4.95	9.43
	200	0.53	1.64	2.19	3.28	4.34	5.37	10.23
	500	0.58	1.80	2.43	3.60	4.79	5.95	11.31
	1000	0.60	1.92	2.57	3.84	5.13	6.35	12.10
p=20	100	0.80	2.44	3.27	4.89	6.46	8.03	15.26
	200	0.84	2.60	3.47	5.20	6.85	8.53	16.24
	500	0.89	2.80	3.73	5.57	7.40	9.22	17.54
	1000	0.92	2.92	3.90	5.86	7.77	9.70	18.54

4.2.4 Comparaison de la carte de contrôle EWMA avec la CCFN

Si dans l'équation (4.6) tous les coefficients a_i sont nuls sauf le coefficient a_0 , alors la fonction de transfert en z du filtre ARMA devient:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{a_0}{1 - \sum_{j=1}^m b_j z^{-j}} \quad (4.17)$$

et en revenant dans le domaine temporel on obtient:

$$y_k = a_0 x_k + b_1 y_{k-1} + b_2 y_{k-2} + \dots + b_m y_{k-m} \quad (4.18)$$

La relation précédente, qui montre que la sortie y_k ne dépend que de la valeur actuelle de l'entrée x et des m valeurs précédentes de la sortie, définit le filtre appelé **autorégressif (AR)** dont le schéma est donné dans la figure IV.6.

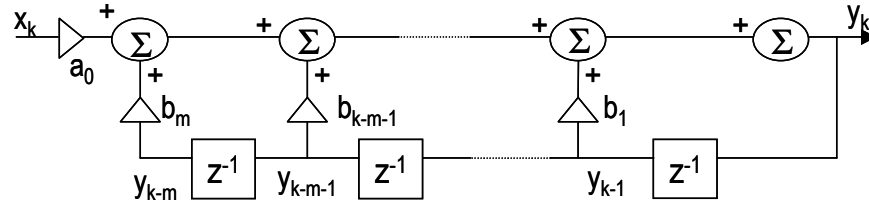


figure IV.6 – Schéma du filtre AR

Il est évident que la statistique EWMA ne représente qu'un cas particulier de filtre (AR). En effet, si dans l'équation (4.18) on considère que tous les coefficients b_j sont nuls pour $j = 2, 3 \dots m$, alors on a:

$$y_k = a_0 x_k + b_1 y_{k-1} \tag{4.19}$$

et comme selon l'équation (4.8) $a_0 = 1 - b_1$, l'équation (4.19) devient:

$$y_k = a_0 x_k + (1 - a_0) y_{k-1} \tag{4.20}$$

On obtient ainsi la statistique EWMA présentée dans l'équation (4.1) et dont le modèle schématique est donné dans la figure IV.7.

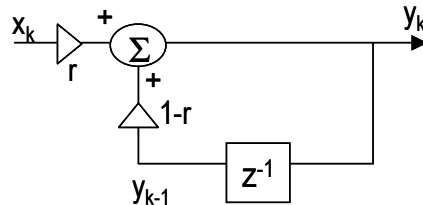


figure IV.7 – Schéma de la statistique EWMA

Pour la construction de la carte de contrôle EWMA on n'a besoin que de la valeur courante à l'entrée x_k et de la valeur précédente de la sortie y_k . Par contre, pour la construction de la CCFN on tient compte de plusieurs valeurs antérieures à l'entrée et à la sortie. Reste à voir si la complexité plus élevée de la carte CCFN au niveau de la conception va se justifier aussi au niveau de ses performances dans la détection des dérèglages faibles.

A partir de l'équation (4.3) et de l'équation (4.11), on peut déduire une équation d'équivalence entre les paramètres de la carte EWMA et CCFN:

$$\sigma_{y_k^{EWMA}}^2 = [1 - (1 - r)^{2k}] \frac{r}{2 - r} \sigma_x^2 = \sigma_{y_k^{CCFN}}^2 = \sigma_x^2 \left(\sum_{i=1}^k c_i^2 \right) \tag{4.21}$$

ou, après les simplifications:

$$[1 - (1-r)^{2k}] \frac{r}{2-r} = \sum_{i=1}^k c_i^2 \quad (4.22)$$

Comme on l'a démontré (tableau 4.2), quand k devient important la somme de la partie droite dans l'équation (4.22) est convergente et de même la partie gauche approche asymptotiquement une valeur constante. On obtient ainsi:

$$\frac{r}{2-r} = C \quad (4.23)$$

où $C = \sum_{i=1}^k c_i^2$ est une constante dépendant de l'ordre et de la fréquence de coupure du filtre. Par conséquent, étant donné un certain filtre passe-bas on peut déterminer à l'aide de l'équation (4.23) le paramètre r de la carte de contrôle EWMA ayant le même écart-type pour la caractéristique affichée sur la carte.

Nous avons comparé les poids attribués aux différentes observations dans le cas de la carte de contrôle EWMA et de la carte de contrôle CCFN utilisant un filtre de type passe-bas Butterworth. Nous avons considéré dans tous les exemples traités qu'un échantillon est prélevé par seconde, d'où l'unité de mesure pour la fréquence en hertz. Cela ne change en rien le principe de la méthode qui peut s'appliquer pour toute fréquence d'échantillonnage.

Dans la figure IV.8.a on observe la décroissance exponentielle de poids typique pour la carte EWMA (ici, $r = 0.05$; $r = 0.15$ et $r = 0.3$), décroissance d'autant plus forte que le paramètre r est proche de 1.

Dans la figure IV.8.b pour la même fréquence de coupure du filtre ($F_c=0.06$ Hz) on peut remarquer l'influence de l'ordre du filtre sur les poids ($L = 2$ à 8). Une première différence par rapport à la carte de contrôle EWMA est observable: l'existence des poids négatifs. En effet, les poids pour la CCFN semble suivre une distribution de la forme d'un sinus amorti. Le sinus s'amortit d'autant plus vite que l'ordre du filtre est réduit. L'observation courante n'a pas le plus grand poids comme c'était le cas pour la carte de contrôle EWMA. Le maximum de poids est donné à un certain nombre de valeurs antérieures qui sont d'autant plus éloignées de l'observation courante que l'ordre du filtre est élevé.

Dans la figure IV.8.c pour le même ordre du filtre on peut remarquer l'influence de la fréquence de coupure sur l'affectation des poids. Si la fréquence de coupure est élevée, un nombre réduit parmi les dernières observations ont un poids important dans le calcul de l'observation courante et les autres n'ont presque aucune influence. Plus on diminue la fréquence de coupure et plus on pondère une séquence plus longue d'observations antérieures mais avec des poids plus réduits.

Finalement dans la figure IV.8.d on peut faire une comparaison entre l'affectation des poids pour le filtre Butterworth ($L = 2$, $F_c = 0.06$ Hz) de type passe-bas et pour la carte EWMA ($r = 0.23$) qui donne à la sortie y la même variance. On observe

que par rapport au modèle EWMA dans le cas de la CCFN le poids associé aux dernières observations (les plus récentes en ordre chronologique) est plus faible.

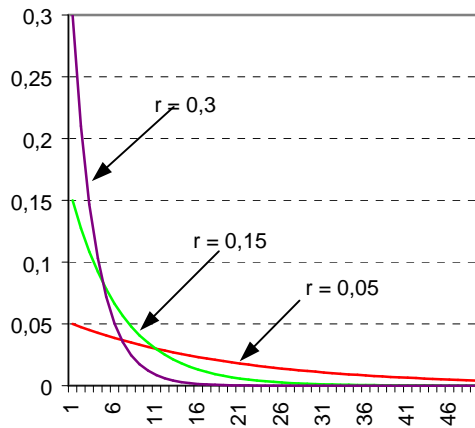


figure IV.8.a – Poids EWMA

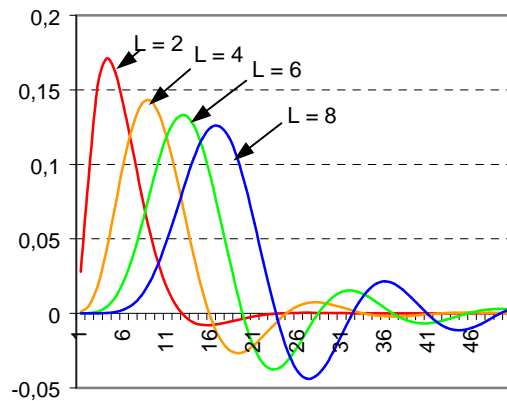


figure IV.8.b – Poids CCFN en fonction de l'ordre du filtre L

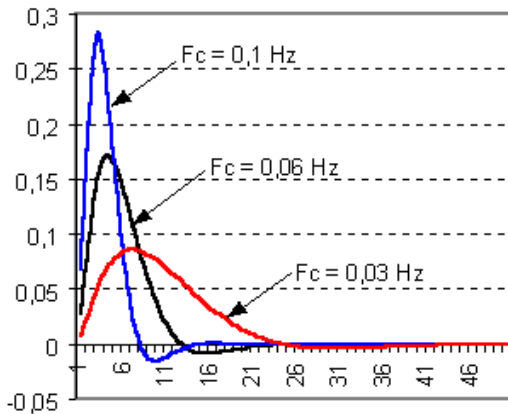


figure IV.8.c – Poids CCFN en fonction de la fréquence de coupure F_c du filtre

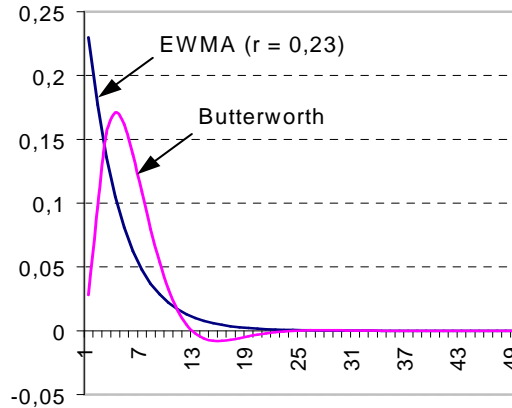


figure IV.8.d – Poids EWMA et CCFN pour la même variance à la sortie

4.2.5 Evaluation de la performance de la carte de contrôle CCFN

Pour déterminer la POM_1 de la carte de contrôle EWMA plusieurs méthodes ont été utilisées. Roberts [103] a été le premier à donner des estimations de la POM_1 de la carte de contrôle EWMA en faisant des simulations.

Crowder [19] a évalué les propriétés de la carte EWMA en résolvant un système d'équations intégrales. La résolution de ce système d'équations repose sur des méthodes approximatives.

Lucas et Saccucci [74, 106] ont utilisé la méthode des chaînes de Markov pour calculer la POM_1 de la carte de contrôle EWMA. Cette méthode a été proposée pour la première fois par Brook et Evans [7] pour le calcul de la POM_1 de la carte de contrôle CUSUM. L'approche des chaînes de Markov consiste à diviser l'intervalle des valeurs possibles pour la caractéristique affichée sur la carte. Utilisant les propriétés des chaînes de Markov de longueur finie, des expressions pour la POM_1 de la carte de contrôle

peuvent être obtenues. Un résumé de la méthode des chaînes de Markov utilisée pour le calcul de la POM_1 de la carte de contrôle EWMA est donné dans l'annexe A.6.

Champ et Rigdon [10] ont comparé l'approche des chaînes de Markov et la méthode de l'équation intégrale et ont trouvé que les résultats obtenus sont similaires en ce qui concerne la précision des calculs. La méthode basée sur l'utilisation des chaînes de Markov commence par l'approximation du problème à résoudre et ensuite la solution exacte est calculée. Par contre la méthode de l'équation intégrale commence avec le problème exact et ensuite une solution approximative est calculée. Ils ont conclu qu'il est plus avantageux d'utiliser la méthode de l'équation intégrale à chaque fois qu'une équation intégrale peut être trouvée.

Dans le cas de la carte de contrôle par filtrage numérique, les méthodes ci-dessus mentionnées ne peuvent pas être utilisées, à part la simulation. Pour démontrer cela, prenons l'exemple le plus simple en considérant $L = m = n = 2$ dans l'équation (4.7). On obtient ainsi:

$$y_k = a_0x_k + a_1x_{k-1} + a_2x_{k-2} + b_1y_{k-1} + b_2y_{k-2} \quad (4.24)$$

c'est-à-dire un modèle où la valeur courante à la sortie y_k dépend de la valeur présente x_k à l'entrée et de deux valeurs antérieures x_{k-1} et x_{k-2} mais aussi de deux valeurs précédentes de la sortie y_{k-1} et y_{k-2} .

Une des hypothèses fondamentales dans le calcul de la POM de la carte de contrôle EWMA par la méthode des chaînes Markov est que les observations x_k , $k = 1, 2, \dots$, sont indépendantes et identiquement distribuées (selon une distribution normale généralement).

Or si on note $z_k = a_0x_k + a_1x_{k-1} + a_2x_{k-2}$, on peut écrire l'équation (4.24) sous la forme: $y_k = z_k + b_1y_{k-1} + b_2y_{k-2}$ et il est évident que les valeurs z_k , $k = 2, 3, \dots$, ne sont pas indépendantes mais autocorrélées. Deuxièmement, dans l'équation (4.24) la valeur courante y_k dépend de deux valeurs antérieures (y_{k-1} et y_{k-2}) et non pas d'une seule comme c'était le cas pour la carte EWMA. Si l'on essaye d'appliquer la méthodologie des chaînes de Markov décrite par Lucas et Saccucci, même si on divise seulement en trois l'intervalle entre les deux limites de contrôle, on n'arrive pas à résoudre le système d'équations ainsi obtenu car on a plus d'inconnues que d'équations.

Il nous semble intéressant de trouver une méthode de calcul mathématique pour estimer l'état courant en fonction de deux ou plusieurs états antérieurs pour un système décrit par une équation de type (4.7). Le calcul à l'aide d'une formule analytique de la POM de la carte de contrôle par filtrage reste un objectif pour de futures recherches.

Nous avons déterminé à partir de 5000 simulations le coefficient h_{CCFN} intervenant dans l'équation (4.12). Ce coefficient a été déterminé (tableau IV.4) pour différentes POM_0 . Les coefficients a_i et b_j intervenant dans l'équation (4.7) sont calculés pour des filtres passe-bas de type Butterworth d'ordre $L = 2$. On remarque que le coefficient h_{CCFN} ne dépend pas de la fréquence de coupure du filtre mais uniquement de la valeur de la POM_0 .

Nous avons également calculé par simulations la POM_1 de la carte de contrôle pour différents filtres numériques de type passe-bas Butterworth d'ordre $L = 2$. Les résultats obtenus sont présentés dans le tableau IV.5 en fonction de la fréquence de

coupure du filtre et pour une $POM_0 = 500$. Nous avons ajouté sur la dernière ligne du tableau les valeurs de la POM_1 pour la carte de contrôle EWMA optimale selon Lucas et Saccucci [74].

tableau IV.4

$POM_0 = 100$	$POM_0 = 370$	$POM_0 = 500$	$POM_0 = 1000$
$h = 2,57$	$h = 2,98$	$h = 3,08$	$h = 3,28$

Nous remarquons que les performances de la CCFN sont généralement inférieures à celle de la carte EWMA et surtout dans le cas des dérèglages de faibles amplitudes (inférieures à 1σ).

tableau IV.5

Amplitude saut (en σ)	0	0,5	1	2	3	4
FC (Hz)						
0,01	500	49,44	21,17	13,00	10,10	8,51
0,03	500	64,04	13,23	6,43	4,78	3,97
0,04	500	68,46	12,62	5,34	3,93	3,23
0,06	500	74,72	13,39	4,29	3,03	2,47
0,08	500	79,43	14,61	3,76	2,59	2,04
0,1	500	84,32	16,30	3,33	2,19	1,77
0,2	500	96,43	24,27	3,01	1,43	1,07
EWMA (optimale)*	500	28,7	10,2	3,51	1,86	1,21

4.2.6 Exemples d'application de la CCFN

Nous avons généré une série de 100 données ayant une distribution normale de moyenne nulle et écart-type $\sigma = 1$. A partir de l'instant 50 nous avons simulé des causes assignables de type saut en échelon et dérèglage en rampe ayant différentes amplitudes. Ensuite nous avons vérifié leur détection sur la carte de contrôle par filtrage et sur la carte de contrôle EWMA.

Saut en échelon d'amplitude 0.5σ

Pour détecter ce type de dérèglage nous avons utilisé une CCFN basée sur un filtre passe-bas Butterworth d'ordre $L = 2$ et ayant une fréquence de coupure égale à $F_c = 0.01\text{Hz}$. Selon les résultats obtenus dans le précédent paragraphe cette carte de contrôle est la plus adéquate pour détecter les dérèglages de faible amplitude. Dans la figure IV.9.a on présente la détection du dérèglage avec la CCFN et dans la figure IV.9.b la détection du même dérèglage avec la carte de contrôle EWMA optimale pour détecter ce type de dérèglage, selon Lucas et Saccucci [74], ayant les paramètres $r = 0.05$ et $h_{EWMA} = 2.616$. Les limites de contrôle sur les deux cartes sont fixées pour assurer une $POM_0 = 500$. Sur la CCFN nous avons détecté le dérèglage après 37 observations et sur la carte de contrôle EWMA après 27 observations.

* Selon Lucas et Saccucci [74]

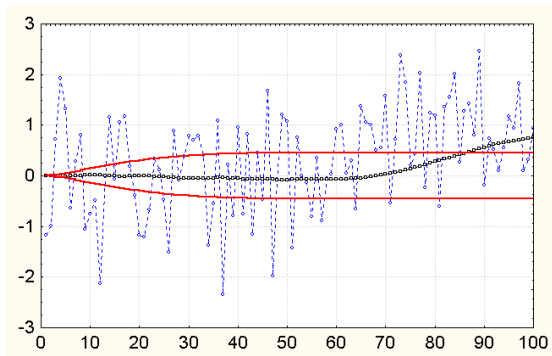


figure IV.9.a - CCFN

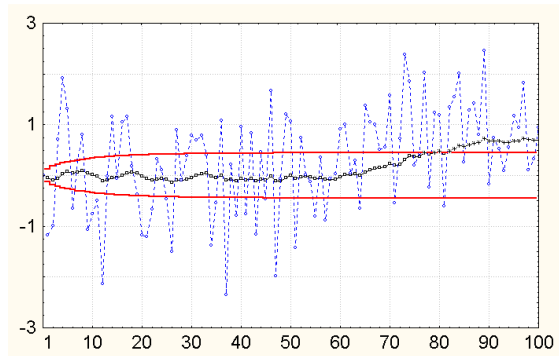


figure IV.9.b - EWMA

Saut en échelon d'amplitude 1σ

Pour détecter ce type de dérèglement nous avons utilisé une CCFN basée sur un filtre passe-bas Butterworth d'ordre $L = 2$ et ayant une fréquence de coupure égale à $F_c = 0.04\text{Hz}$. La carte de contrôle EWMA optimale pour la détection de ce type de dérèglement a les paramètres suivants: $r = 0.12 - 0.15$ et $h_{EWMA} = 2.858 - 2.907$. Dans la figure IV.10.a on présente la détection du dérèglement avec la CCFN et dans la figure IV.10.b la détection du même dérèglement avec la carte de contrôle EWMA.

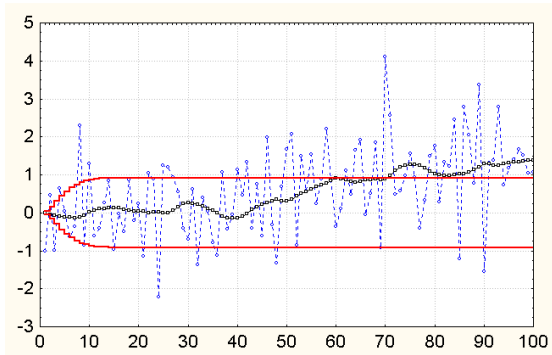


figure IV.10.a - CCFN

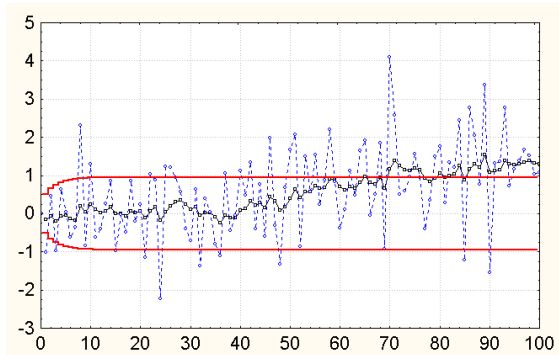


figure IV.10.b - EWMA

Dans ce cas les deux cartes présentent des propriétés très similaires en détection ce que l'on peut constater aussi dans le tableau IV.5. Le dérèglement est détecté après 20 observations dans les deux cas.

Saut en échelon d'amplitude 1.5σ

Pour détecter ce type de dérèglement nous avons utilisé une CCFN basée sur un filtre passe-bas Butterworth d'ordre $L = 2$ et ayant une fréquence de coupure égale à $F_c = 0.06\text{Hz}$. La carte de contrôle EWMA optimale pour la détection de ce type de dérèglement a les paramètres suivants: $r = 0.25$ et $h_{EWMA} = 2.998$. Dans la figure IV.11.a on présente la détection du dérèglement avec la CCFN et dans la figure IV.11.b la détection du même dérèglement avec la carte de contrôle EWMA. Les deux cartes de contrôle détectent le dérèglement après 6 observations.

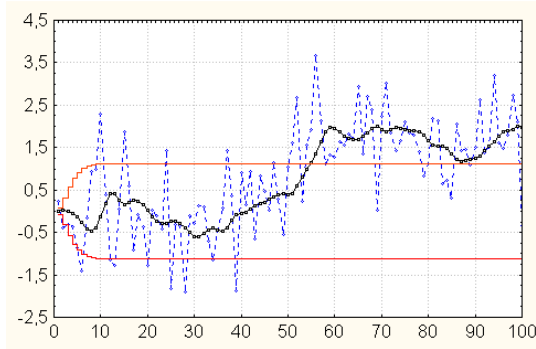


figure IV.11.a - CCFN

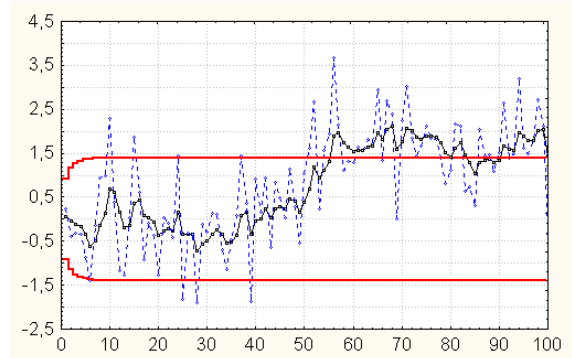


figure IV.11.b - EWMA

Saut en rampe d'amplitude 1.5σ

La carte de contrôle basée sur un filtre passe-bas de type Butterworth d'ordre $L=2$ et ayant une fréquence de coupure $F_c = 0.1\text{Hz}$ détecte le dérèglement (figure IV.12.a) après 21 observations. Pourtant, un très grand nombre d'observations ultérieures à l'observation qui dépasse la limite supérieure de contrôle sont situées sous la limite de contrôle. Donc, si l'opérateur ne prend pas une décision tout de suite il risque de ne pas intervenir car les prochains points sont situés entre les deux limites de contrôle. Ainsi, globalement la détection du dérèglement est retardée.

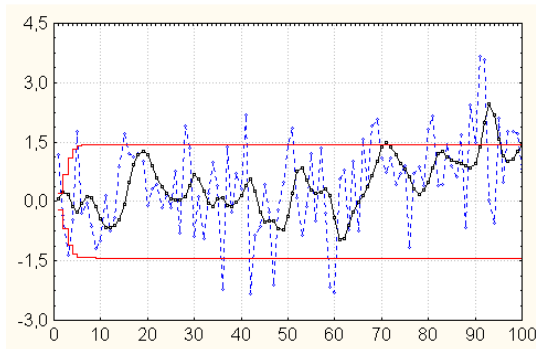


figure IV.12.a - CCFN ($F_c = 0.1\text{Hz}$)

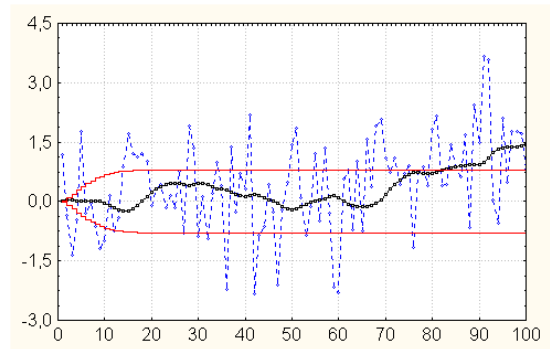


figure IV.12.b - CCFN ($F_c = 0.03\text{ Hz}$)

Dans la figure IV.12.b nous présentons, pour le même cas, la détection utilisant une carte de contrôle par filtrage basée sur un filtre passe-bas Butterworth d'ordre $L=2$ avec une fréquence de coupure $F_c = 0.03\text{ Hz}$. Le dérèglement est détecté après 32 observations, mais une fois que la limite supérieure de contrôle a été dépassée, il ne reste plus de doute concernant le dérèglement. Nous recommandons dans le cas des dérèglages en rampe de faible amplitude l'utilisation des cartes de contrôles basées sur des filtres passe-bas ayant une fréquence de coupure basse qui ont l'avantage de mettre clairement en évidence la tendance de la moyenne.

4.3 Carte de contrôle spectrale

4.3.1 Introduction

Nous allons d'abord présenter dans cette introduction quelques notions théoriques concernant la transformée de Fourier en discret qui nous serviront, par la suite, pour la définition du principe de construction de la carte de contrôle spectrale proposée dans ce paragraphe.

Par analogie avec la transformée de Fourier en continu, dans le cas discret, un signal périodique quelconque $x = \{x_k \mid k = 0, \dots, N-1\}$ de N échantillons peut être entièrement reconstitué par la sommation de $N/2 + 1$ fonctions cosinus et $N/2 + 1$ fonctions sinus ayant les amplitudes convenablement choisies (voir l'équation 4.25). Les amplitudes des sinus et des cosinus sont stockées dans les matrices $A []$ et $B []$.

$$x_k = \sum_{n=0}^{N/2} A[n] \cos\left(2\pi k \frac{n}{N}\right) + \sum_{n=0}^{N/2} B[n] \sin\left(2\pi k \frac{n}{N}\right), \quad k = 0, \dots, N-1 \quad (4.25)$$

L'équation (4.25) nous montre que les fonctions $\cos\left(2\pi k \frac{n}{N}\right)$ et $\sin\left(2\pi k \frac{n}{N}\right)$ forment une base orthogonale pour le signal x . En choisissant d'une manière adéquate les coefficients $A []$ et $B []$ on peut reconstituer à l'aide de cette base n'importe quel signal.

Donc, la transformée de Fourier en discret (figure IV.13), transforme un signal d'entrée x de N points en deux séries de $N/2+1$ points à la sortie contenant les amplitudes (coefficients) des sinus et des cosinus. Le domaine fréquentiel contient exactement la même information que le domaine temporel mais sous une forme différente. Si un des deux domaines est connu, alors on peut déduire facilement l'autre.

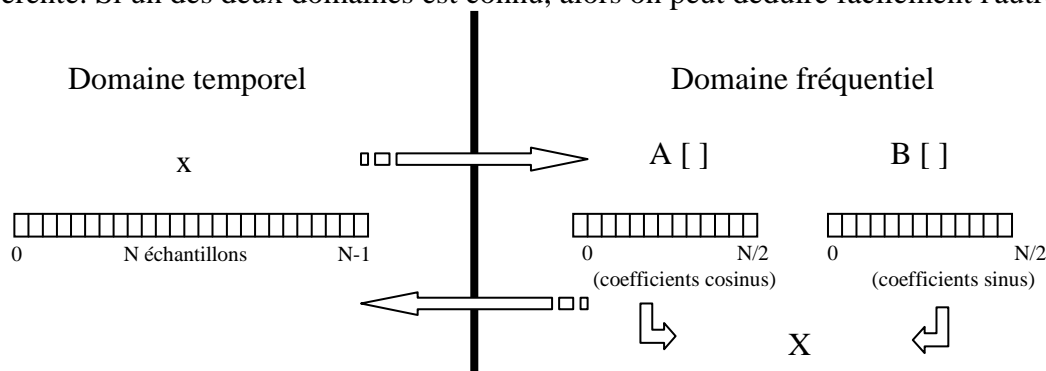


figure IV.13 - la transformée de Fourier en discret

Pour passer du domaine temporel au domaine fréquentiel (transformée de Fourier directe) on doit calculer les deux matrices $A []$ et $B []$ à l'aide des équations (4.26).

$$\begin{aligned}
 A[0] &= \frac{1}{N} \sum_{k=0}^{N-1} x_k; & A\left[\frac{N}{2}\right] &= \frac{1}{N} \sum_{k=0}^{N-1} x_k \cos(k\pi) = \frac{1}{N} \sum_{k=0}^{N-1} (-1)^k x_k \\
 A[n] &= \frac{2}{N} \sum_{k=0}^{N-1} x_k \cos\left(2\pi k \frac{n}{N}\right), & \text{pour } n = 1, \dots, \frac{N}{2} - 1 \\
 B[n] &= \frac{2}{N} \sum_{k=0}^{N-1} x_k \sin\left(2\pi k \frac{n}{N}\right), & \text{pour } n = 0, \dots, \frac{N}{2}
 \end{aligned} \tag{4.26}$$

Réciproquement, pour passer du domaine des fréquences au domaine du temps on utilise la transformée de Fourier inverse (voir l'équation 4.25).

Par la transformation des coordonnées cartésiennes en coordonnées polaires: $A \cos(x) + B \sin(x) = M \cos(x + \varphi)$, l'équation (4.25) peut s'écrire de la manière suivante:

$$x_k = \sum_{n=0}^{N/2} M[n] \cos\left(2\pi k \frac{n}{N} + \Phi[n]\right), \quad k = 0, \dots, N-1 \tag{4.27}$$

où

$$M[n] = \sqrt{A^2[n] + B^2[n]} \tag{4.28}$$

$$\Phi[n] = \arctan\left(\frac{B[n]}{A[n]}\right)$$

sont appelés le module (ou l'amplitude) et la phase de la transformée de Fourier discrète.

Connaissant le module $M[n]$ et la phase $\Phi[n]$, les coordonnées cartésiennes peuvent être calculées à l'aide du système d'équations:

$$\begin{aligned}
 A[n] &= M[n] \cos(\Phi[n]) \\
 B[n] &= M[n] \sin(\Phi[n])
 \end{aligned} \tag{4.29}$$

4.3.2 Principe de construction de la carte de contrôle spectrale

Nous allons développer l'équation (4.25) en tenant compte du fait que pour $n = 0$ et pour $n = N/2$ les coefficients $B[n]$ sont nuls (voir l'équation 4.26):

$$\begin{aligned}
 x_k &= A[0] + A\left[\frac{N}{2}\right] \cos(k\pi) + \sum_{n=1}^{\frac{N}{2}-1} A[n] \cos\left(2\pi k \frac{n}{N}\right) + \sum_{n=1}^{\frac{N}{2}-1} B[n] \sin\left(2\pi k \frac{n}{N}\right) \\
 x_k &= A[0] + A\left[\frac{N}{2}\right] \cos(k\pi) + \sum_{n=1}^{\frac{N}{2}-1} \left\{ A[n] \cos\left(2\pi k \frac{n}{N}\right) + B[n] \sin\left(2\pi k \frac{n}{N}\right) \right\} \\
 x_k &= A[0] + A\left[\frac{N}{2}\right] \cos(k\pi) + \sum_{n=1}^{\frac{N}{2}-1} M[n] \cos\left(2\pi k \frac{n}{N} + \Phi[n]\right)
 \end{aligned} \tag{4.30}$$

Supposons maintenant, que x représente un bruit blanc. En calculant la transformée de Fourier de x , nous allons obtenir, selon les équations (4.26) et le théorème central limite, des coefficients $A[]$ et $B[]$ qui sont des variables aléatoires indépendantes et distribuées selon une loi normale. La distribution de deux variables aléatoires $A[]$ et $B[]$ s'approche de la distribution normale quelle que soit la distribution du bruit. Elle est d'autant plus proche de la distribution normale que N est grand.

Notons \bar{A} la moyenne et σ_A l'écart-type de la distribution de la variable aléatoire $A^*[]$ et \bar{B} la moyenne et σ_B l'écart-type de la distribution de la variable aléatoire $B^*[]$, avec: $A^*[] = \{A[n] | n = 1, \dots, N/2 - 1\}$ et $B^*[] = \{B[n] | n = 1, \dots, N/2 - 1\}$ les matrices des coefficients des fonctions cosinus et sinus dans l'expression de la transformée de Fourier de $x = \{x_k | k = 0, \dots, N-1\}$ après avoir éliminé les premiers ($A[0]$ et $B[0]$) et les derniers ($A[N/2]$ et $B[N/2]$). On a donc:

$$\begin{aligned} A^*[] &\sim N(\bar{A}, \sigma_A) \\ B^*[] &\sim N(\bar{B}, \sigma_B) \end{aligned} \quad (4.31)$$

autrement dit:

$$\begin{aligned} A_{CR}^*[] &= \frac{A^*[] - \bar{A}}{\sigma_A} \sim N(0,1) \\ B_{CR}^*[] &= \frac{B^*[] - \bar{B}}{\sigma_B} \sim N(0,1) \end{aligned} \quad (4.32)$$

De la même manière (tenant compte de l'équation 4.28) on obtient:

$$M_{CR}^2[] = A_{CR}^{*2}[] + B_{CR}^{*2}[] = \left(\frac{A^*[] - \bar{A}}{\sigma_A} \right)^2 + \left(\frac{B^*[] - \bar{B}}{\sigma_B} \right)^2 \quad (4.33)$$

qui, étant la somme de deux variables aléatoires ayant une distribution normale de moyenne nulle et l'écart type égal à 1, sera une variable aléatoire ayant une distribution de khi-deux* avec un nombre de degrés de liberté égal à 2:

$$M_{CR}^2[] \sim \chi_{(2ddl)}^2 \quad (4.34)$$

Nous proposons une nouvelle carte de contrôle basée sur l'utilisation de $M_{CR}^2[]$ nommée "Carte de contrôle spectrale". Sur cette carte, nous représentons les valeurs

* La fonction de densité de probabilité de la distribution de khi-deux ayant n degrés de liberté est:

$$f(x) = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} x^{n/2-1} e^{-x/2}, \quad x > 0$$

La moyenne de la distribution est $\mu = n$ et la variance est $\sigma^2 = 2n$

$\{ M_{CR}^2 [n] \mid \text{pour } n=1, \dots, N/2-1 \}$ et la limite de contrôle sur la carte sera fixée à $LC = \chi_{\alpha}^2$, où χ_{α}^2 représente le quantile d'ordre α de la distribution de khi-deux. On donne dans le tableau IV.6 les différentes valeurs pour la limite de contrôle de la carte en fonction de la POM_0 (entre les crochets nous avons donné le risque α) souhaitée.

tableau IV.6

POM_0	100	200	500	1000	2000
$[\alpha (\%)]$	[1]	[0,5]	[0,2]	[0,1]	[0,05]
LC	9,21	10,60	12,43	13,82	17,03

Les questions qui se posent maintenant: que se passe-t-il dans le cas de la présence de causes assignables de type saut en échelon ou dérive en rampe? Comment va-t-on détecter ces causes assignables sur la carte spectrale? Quel doit-être le nombre d'échantillons N (taille de la fenêtre temporelle) pour que l'on dispose d'une bonne résolution dans le domaine spectral?

4.3.3 Etude du spectre des causes assignables

La carte de contrôle spectrale, repose sur le fait que la transformé de Fourier est additive. Cela veut dire que, étant donnés deux signaux $x = \{x_k \mid k = 0, \dots, N-1\}$ et $y = \{y_k \mid k = 0, \dots, N-1\}$ ayant respectivement les transformées de Fourier $X = \{Ax []; Bx []\}$ et $Y = \{Ay []; By []\}$, le signal $z = x + y$ aura la transformée de Fourier $Z = \{Az []; Bz []\}$ où $Az [] = Ax [] + Ay []$ et $Bz [] = Bx [] + By []$.

Par conséquent, un changement dans le spectre du bruit du à l'apparition d'une cause assignable, devra se manifester dans les coefficients $A []$ et $B []$ de la décomposition spectrale et donc sur la carte de contrôle spectrale M_{CR}^2 .

Mais avant de montrer comment est faite la détection dans le domaine spectral nous allons analyser plus en détail la composition spectrale de deux causes assignables: le saut en échelon et la dérive en rampe.

Le spectre du saut en échelon

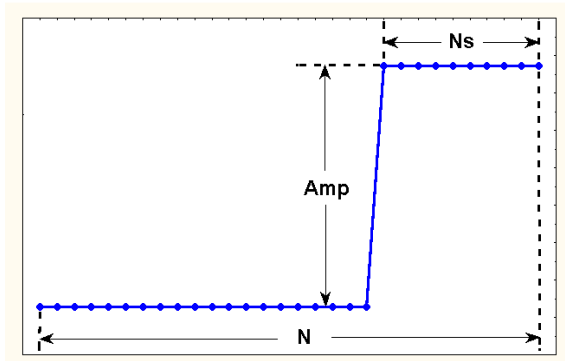
Nous avons analysé le spectre du saut en échelon en fonction de plusieurs paramètres (voir figure IV.14.a):

- ✓ le nombre total d'échantillons: N
- ✓ le nombre d'échantillons correspondant au saut: N_s
- ✓ l'amplitude du saut: Amp

Dans la figure IV.14.b et dans la figure IV.14.c on peut voir l'évolution des coefficients cosinus ($A^* []$) et sinus ($B^* []$) utilisés par la carte spectrale M_{CR}^2 en fonction de N pour un $N_s = 10$ et $Amp = 1$. On constate que les coefficients cosinus et sinus décrivent une courbe sinusoïdale amortie. Si N est grand, l'amplitude de la première sinusoïde est réduite. Au fur et à mesure que l'on réduit N l'amplitude de la première sinusoïde augmente.

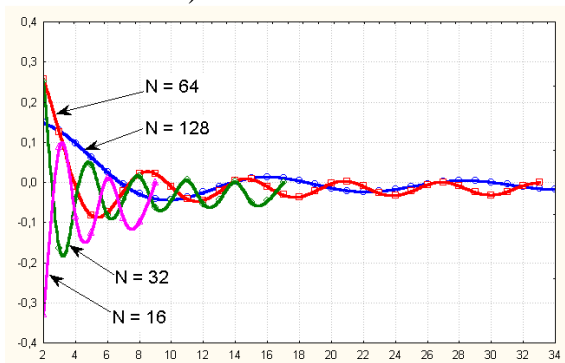
Dans la figure IV.14.d et IV.14.e on remarque que dans le cas où N et N_s sont constants, l'amplitude de la sinusoïde amortie est proportionnelle à l'amplitude du saut.

Plus l'amplitude du saut est grande et plus la hauteur de la sinusoïde principale est grande.

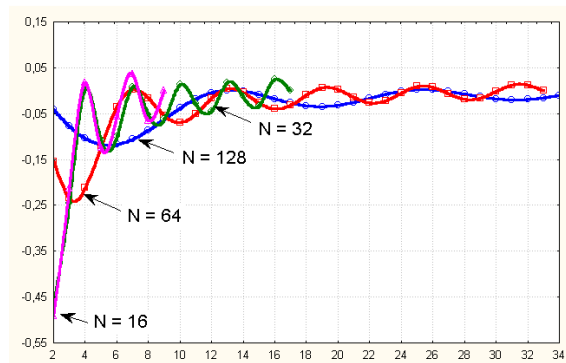


a) saut en échelon

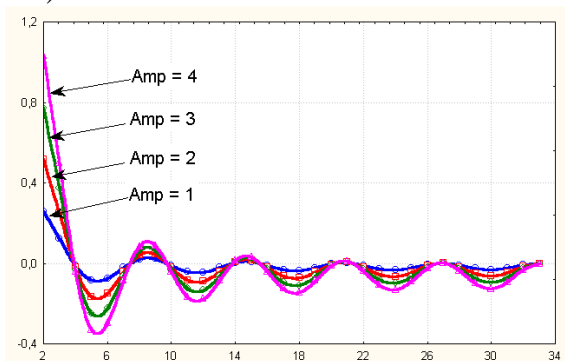
figure IV.14 - Analyse des composantes spectrales du saut en échelon



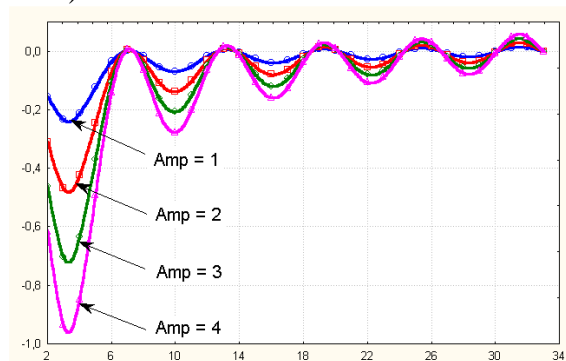
b) coefficients cosinus en fonction de N



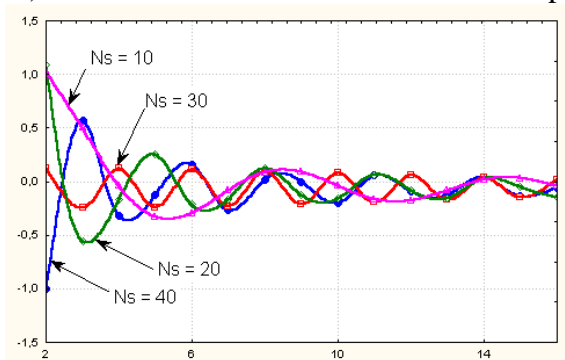
c) coefficients sinus en fonction de N



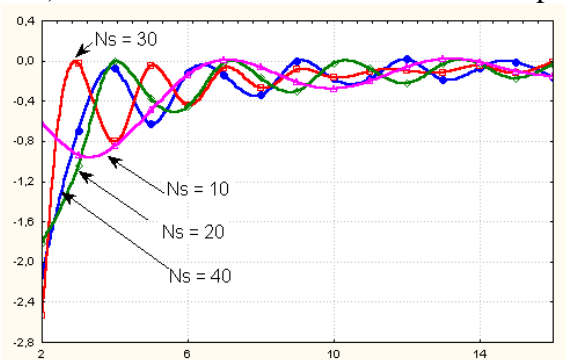
d) coefficients cosinus en fonction de Amp



e) coefficients sinus en fonction de Amp



f) coefficients cosinus en fonction de Ns



g) coefficients sinus en fonction de Ns

Finalement, dans la figure IV.14.f et IV.14.g, on peut constater l'évolution des coefficients cosinus et sinus en fonction de Ns dans le cas où N et Amp sont constants.

Le spectre de la dérive en rampe

En modifiant les mêmes paramètres cités précédemment, nous avons analysé le spectre de la dérive en rampe.

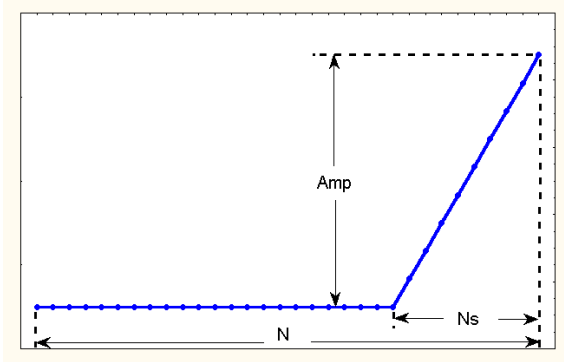
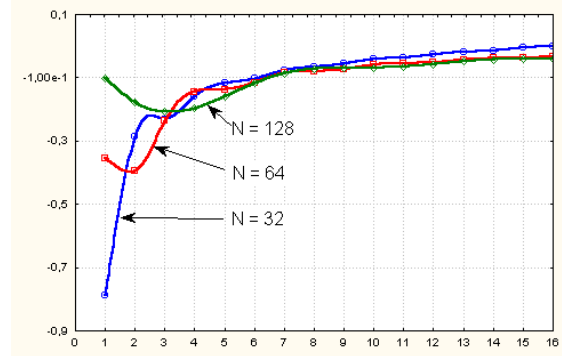
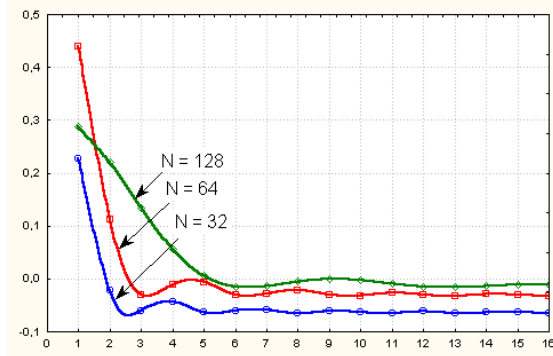


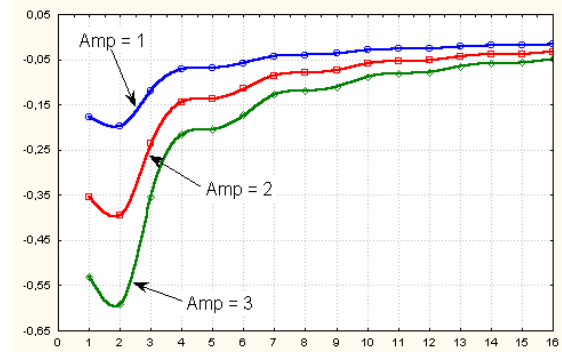
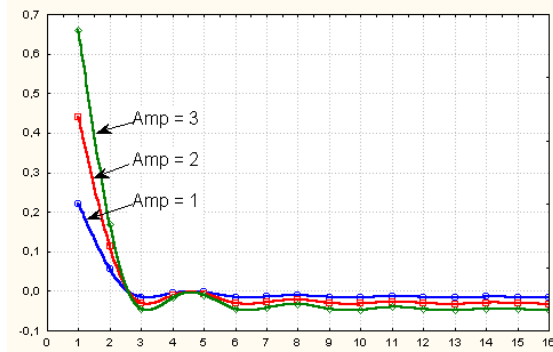
figure IV.15 - Analyse des composantes spectrales de la dérive en rampe

a) dérive en rampe



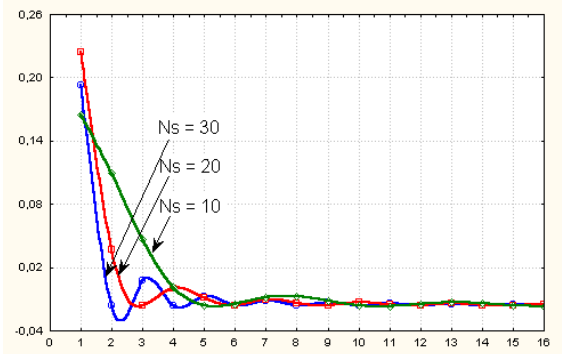
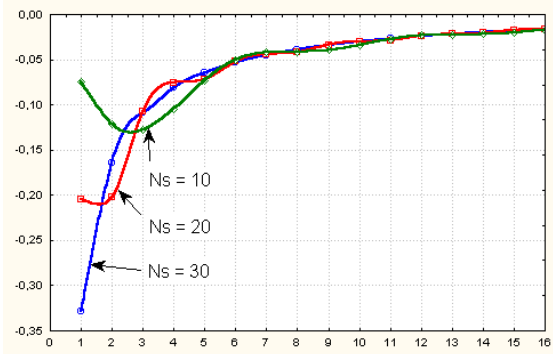
b) coefficients cosinus en fonction de N

c) coefficients sinus en fonction de N



d) coefficients cosinus en fonction de Amp

e) coefficients sinus en fonction de Amp



f) coefficients cosinus en fonction de Ns

g) coefficients sinus en fonction de Ns

Les résultats, représentant l'évolution des coefficients cosinus et sinus $A^*[]$ et $B^*[]$ utilisés par la carte spectrale M_{CR}^2 sont donnés dans la figure IV.15. Les mêmes remarques que pour le saut en échelon restent valables dans le cas de la dérive en rampe.

Il a été démontré (voir le paragraphe 4.3.2) que les coefficients cosinus et sinus du bruit blanc sont des variables aléatoires indépendantes et identiquement distribuées. Donc, on peut les assimiler à un bruit blanc à leur tour. On a également remarqué que leur distribution s'approche de la distribution normale d'autant plus que le nombre d'échantillons est grand.

Nous avons aussi étudié l'évolution des coefficients cosinus $A^*[]$ et sinus $B^*[]$ utilisés par la carte spectrale M_{CR}^2 pour deux causes assignables: le saut en échelon et la dérive en rampe. Quand la cause assignable est noyée dans le bruit, la transformée de Fourier du signal ainsi obtenu aura les coefficients cosinus et sinus égaux à la somme des coefficients cosinus et des coefficients sinus du bruit et de la cause assignable.

Tenant compte que ces coefficients seront élevés au carré (voir l'équation 4.33), en regardant les figures IV.14 et IV.15, nous pouvons tirer les conclusions suivantes:

- ✓ la cause assignable est d'autant plus facile à détecter que son amplitude est grande;
- ✓ la cause assignable a plus de chances d'être détectée si N (la taille de la fenêtre temporelle) est petit;
- ✓ la présence de la cause assignable sera d'autant plus évidente que N est grand.

4.3.4 Taille de la fenêtre temporelle

Comme la détection des causes assignables se fait dans le domaine spectral, il est souhaitable d'avoir une bonne résolution spectrale. Sachant que pour une variable discrète $x = \{x_k \mid k = 0, \dots, N-1\}$ dans le domaine temporel, sur la carte de contrôle M_{CR}^2 on va représenter $N/2 - 1$ points, il est nécessaire que N soit grand. Autrement dit, si on veut une bonne résolution en fréquence, il faut analyser le signal sur une longue période de temps. Or nous avons vu que la détection de la cause assignable de type saut en échelon ou dérive en rampe est facilitée si la taille de la fenêtre temporelle est réduite. Si on utilise, pour la détection, un nombre réduit d'échantillons du signal, alors la résolution spectrale sera faible mais on gagne en précision dans la détection temporelle. Il existe une méthode simple pour améliorer la résolution spectrale dans le cas où N est petit. La méthode consiste à ajouter à la fin du signal un nombre N' de zéros, ce qui ne change rigoureusement rien à la valeur de la transformée de Fourier. Par contre, la résolution spectrale sera ainsi augmentée car, après l'ajout des zéros, le nombre total des points dans le domaine du temps devient $N + N'$ et donc sur la carte de contrôle spectrale M_{CR}^2 au lieu d'avoir seulement $N/2 - 1$ points on va avoir $(N+N')/2 - 1$ points. Nous recommandons l'utilisation de cette technique surtout quand la taille de la fenêtre d'analyse temporelle du signal est réduite. Nous allons illustrer cet aspect pour la détection d'un saut en échelon et pour la détection d'une dérive en rampe ci-dessous.

Saut en échelon

Dans la figure IV.16.a, nous présentons le cas d'une variable aléatoire $x = \{x_k \mid k = 1, \dots, 64\}$ ayant une distribution normale de moyenne nulle et écart-type $\sigma = 1$, dont la moyenne a subi un saut en échelon d'amplitude 1σ à partir de l'observation $k = 41$.

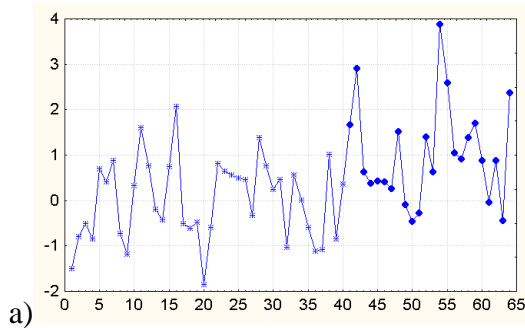
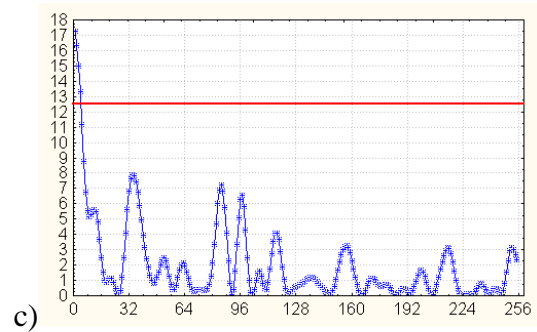
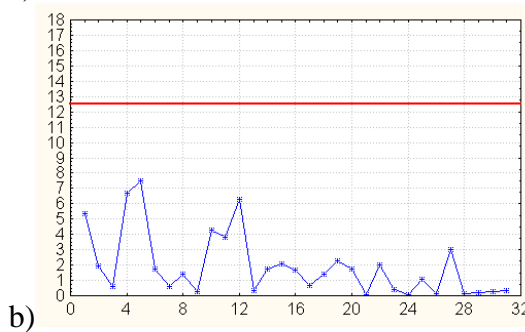


figure IV.16

- a) saut en échelon d'amplitude 1σ dans le domaine du temps
- b) carte de contrôle spectrale
- c) carte de contrôle spectrale de haute résolution



Dans la figure IV.16.b nous avons la carte de contrôle M_{CR}^2 avec la limite de contrôle fixée à 12,43 pour garantir une POM_0 de 500. On remarque qu'aucun point sur la carte de contrôle ne dépasse la limite de contrôle. Cela est dû au fait que la résolution spectrale est faible ($64/2 - 1 = 31$ points seulement sur la carte) et donc la cause assignable passe inaperçue. Pour remédier à cet inconvénient nous avons artificiellement ajouté 448 zéros à la fin de la séquence de 64 valeurs de x . De cette manière, nous n'avons pas changé la transformée de Fourier du signal mais nous avons augmenté la résolution spectrale d'un facteur important. En ajoutant les 448 zéros, le nombre total des points dans le domaine du temps devient $448 + 64 = 512$. Nous avons choisi d'ajouter 448 valeurs pour que le nombre total d'observations soit une puissance de 2. Ainsi, pour le calcul de la composition spectrale du signal, on peut utiliser un algorithme de calcul de la transformée de Fourier nommé la transformée de Fourier rapide, nécessitant moins de temps de calcul.

Après l'ajout des zéros, sur la carte de contrôle spectrale M_{CR}^2 nous obtenons un nombre $N = 512/2 - 1 = 255$ de points au lieu de 31 points et donc la résolution spectrale se trouve multipliée par un facteur d'environ 8. Dans la figure IV.16.c nous présentons la carte de contrôle spectrale M_{CR}^2 pour cette situation. Nous pouvons remarquer que la limite de contrôle a été dépassée. On remarque également que l'allure du spectre ne change pas mais que c'est la résolution qui s'est améliorée. Dans le cas d'une cause assignable de type saut en échelon, la limite de contrôle sur la carte spectrale est franchie toujours en basse-fréquence c'est-à-dire dans la partie gauche sur la carte, près de l'origine.

Déréglage en rampe

Dans la figure IV.17.a, nous présentons le cas d'une variable aléatoire $x = \{x_k \mid k = 1, \dots, 64\}$ ayant une distribution normale de moyenne nulle et l'écart-type $\sigma = 1$, dont la moyenne a subi un saut en rampe d'amplitude 1σ à partir de l'observation $k = 41$.

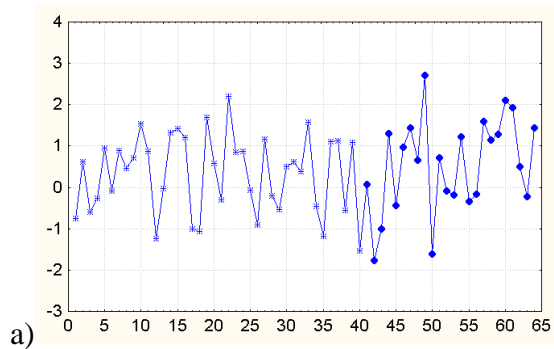
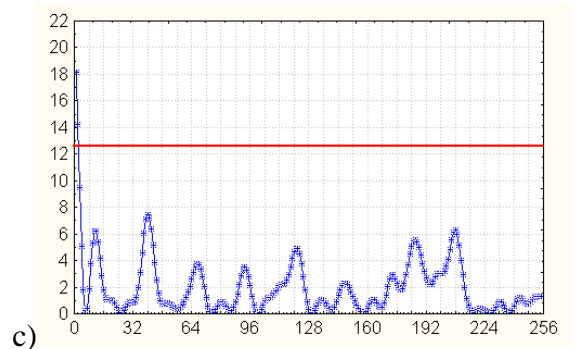
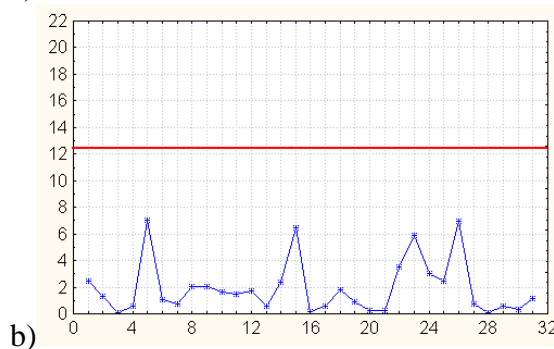


figure IV.17

- a) saut en échelon d'amplitude 1σ dans le domaine du temps
- b) carte de contrôle spectrale
- c) carte de contrôle spectrale de haute résolution



Dans la figure IV.17.b nous avons la carte de contrôle M_{CR}^2 avec la limite de contrôle fixée à 12,43 pour garantir une POM_0 de 500. On remarque qu'aucun point sur la carte de contrôle ne dépasse la limite de contrôle à cause de la faible résolution spectrale. Dans la figure IV.17.c nous présentons la carte de contrôle spectrale M_{CR}^2 après l'ajout de 448 zéros. Nous pouvons remarquer que la limite de contrôle a été dépassée.

4.3.5 Détection des causes assignables avec la carte de contrôle spectrale

Pour la détection des causes assignables de type saut en échelon ou dérive en rampe nous proposons l'utilisation de la carte spectrale M_{CR}^2 (figure IV.18). Elle consiste à faire passer une fenêtre glissante d'une certaine taille à travers le signal et à analyser en permanence le contenu spectral de cette fenêtre.

Dans cet exemple (figure IV.18.a), nous avons simulé une variable aléatoire ayant une distribution normale de moyenne nulle et d'écart-type $\sigma = 1$. A partir de l'observation 32 et jusqu'à l'observation 48 nous avons simulé un déréglage de type saut en échelon d'amplitude 1σ . Nous avons aussi simulé une dérive en rampe d'amplitude 2σ entre les observations 65 et 96.

Sachant que ces deux types de déréglements sont plus facilement détectables si la taille de la fenêtre temporelle est réduite nous avons choisi d'utiliser une fenêtre glissante de 16 points.

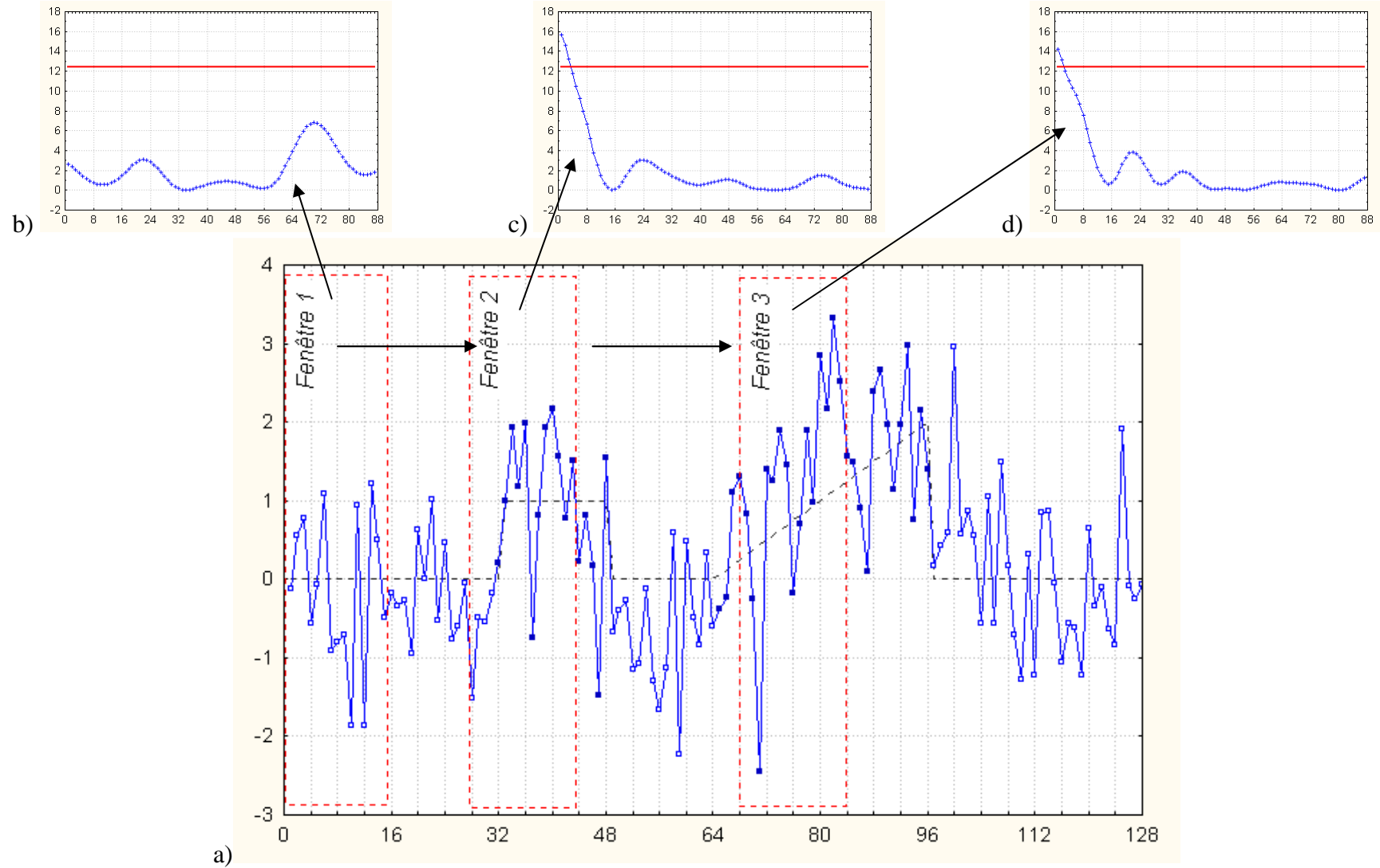


figure IV.18 – Détection des dérèglages à l'aide de la carte de contrôle spectrale

Nous avons ajouté 160 zéros à la séquence de points de la fenêtre afin d'améliorer la résolution spectrale de la fenêtre.

Sur la carte de contrôle spectrale nous avons fixé la limite de contrôle à 12,43 pour assurer une POM_0 de 500.

Dans la figure IV.18.b nous avons représenté l'information affichée sur la carte spectrale M_{CR}^2 dans le cas où il n'y a pas de cause assignable présente (position 1 de la fenêtre glissante). On peut remarquer qu'aucun point sur la carte ne dépasse la limite de contrôle.

Dans la figure IV.18.c on donne le contenu de la carte de contrôle spectrale pour la position 2 de la fenêtre glissante correspondant aux observations comprises entre 28 et 44 c'est-à-dire 12 observations après l'apparition du dérèglement saut en échelon. On remarque qu'il y a des points qui dépassent la limite de contrôle signalant ainsi la présence de la cause assignable.

Dans la figure IV.18.d nous présentons la détection du dérèglement en rampe de la moyenne sur la carte de contrôle M_{CR}^2 . La détection a eu lieu pour la position 3 de la fenêtre glissante correspondant aux observations comprises entre 68 et 86, c'est-à-dire 22 observations après le moment d'apparition du dérèglement.

4.3.6 Analyse de l'efficacité de la carte de contrôle spectrale

Nous avons déterminé, à partir de 100 simulations, la POM_1 de la carte de contrôle spectrale pour la détection des dérèglements de type saut en escalier des différentes amplitudes entre $0,75\sigma$ et 3σ . Nous avons choisi une fenêtre temporelle de 16 observations et nous avons ajouté 160 zéros afin d'améliorer la résolution spectrale. Les résultats obtenus sont présentés dans le tableau IV.7.

tableau IV.7

Amp. [σ] POM ₀	0,75	1	1,5	2	2,5	3
100	6,19	5,72	3,13	2	1,73	1,57
200	8,15	7,17	4,17	2,55	2,17	1,92
500	9,92	8,57	5,28	3,27	2,75	2,38
1000	10,81	9,85	5,86	3,79	3,17	2,59
2000	12,17	11,26	7,16	5,03	3,86	3,15
EWMA (POM ₀ =500)	15,9	10,2	5,46	3,51	2,50	1,86

Sur la dernière ligne du tableau IV.7 nous présentons, à titre comparatif, la POM_1 de la carte de contrôle EWMA optimale pour la détection de ces dérèglements (selon Lucas et Saccucci [4]).

On peut remarquer que la carte de contrôle spectrale a de meilleures performances que la carte EWMA, surtout dans le cas où l'amplitude du dérèglement est faible.

4.4 Conclusion

Dans ce chapitre nous avons proposé deux nouvelles cartes de contrôle - la carte de contrôle par filtrage numérique (CCFN) dont le principe de construction repose sur l'utilisation des équations récurrentes définissant les filtres numériques RII (Réponse Impulsionnelle Infinie) ou ARMA (AutoRegressive Moving Average) et la carte de contrôle spectrale (CCS) dont le principe de construction repose sur l'analyse spectrale du signal.

Par sa construction, la carte de contrôle CCFN prend en compte non seulement l'observation courante et l'observation précédente comme c'est le cas de la carte de contrôle EWMA, mais plusieurs observations antérieures. Nous avons utilisé dans les formules récurrentes, caractérisant ces modèles, des coefficients calculés pour définir des filtres passe-bas, et notamment des filtres passe-bas de type Butterworth d'ordre $L=2$. Nous avons décidé de choisir les filtres passe-bas pour réduire la variabilité de la caractéristique mesurée. Nous avons comparé les principes de construction et les performances de deux cartes de contrôle: CCFN et EWMA. Nous avons calculé les valeurs des limites de contrôle et l'efficacité de la carte CCFN à l'aide de simulations. Nous avons montré que certaines méthodes utilisées habituellement pour le calcul des performances des cartes de contrôle "classiques" ne peuvent pas être utilisées pour évaluer la POM de la carte CCFN. Des recherches devront être menées afin de trouver une expression analytique de calcul pour ces valeurs. Par des simulations, nous avons déterminé que les performances de la carte CCFN sont "légèrement" inférieures à celles de la carte EWMA surtout dans le cas des dérèglages de très faibles amplitudes (inférieures ou égales à 1σ). La carte CCFN présente quand même quelques avantages. Ainsi, même si elle détecte le dérèglement plus tard que la carte EWMA, elle présente une caractéristique plus lisse. Cet aspect peut être utile quand les points sur la carte oscillent autour de la limite de contrôle (voir figure IV.10.b). Souvent dans ce genre de situations les opérateurs ne savent pas quelle décision prendre et ils attendent parfois à tort jusqu'à ce que la situation se stabilise d'une certaine façon. Sur la carte de contrôle par filtrage les points ne présentent pas d'oscillations en dents de scie et les tendances sont plus claires. Un deuxième avantage de la carte CCFN est dans le cas des dérèglages en rampe (représentant des dérives lentes comme, par exemple, l'usure d'un outil). Dans ces situations, par l'utilisation des filtres passe-bas ayant une faible fréquence de coupure, on peut détecter la dérive avant même que la limite de contrôle ne soit dépassée, par une inspection visuelle de la caractéristique affichée sur la carte.

Le principe de contrôle de la carte de contrôle spectrale (CCS) est basé sur l'analogie temps-fréquence: une fenêtre glissante est déplacée d'une manière séquentielle dans le temps et le contenu spectral de la fenêtre est analysé. Nous avons montré qu'il est possible d'utiliser l'information contenue dans les coefficients de la décomposition en série de Fourier d'un signal, afin de détecter les changements (sauts en échelon ou dérives en rampe) de la moyenne. La formule analytique de la limite de contrôle sur la carte a été précisée en partant des considérations théoriques concernant la distribution des coefficients dans le spectre d'amplitude du signal. Des exemples d'utilisation de la carte de contrôle pour la détection des dérèglages ont été présentés. Les performances de la carte dans la détection des différents dérèglages ont été calculées à partir des simulations et comparées avec celles de la carte EWMA. Nous avons constaté que la carte de contrôle spectrale présente de très bonnes performances dans la détection des dérèglages faibles voir même supérieures à celles de la carte EWMA notamment dans le cas des dérèglages de faibles amplitudes.

5. Conclusion générale et perspectives

Dans le contexte économique actuel de globalisation des marchés, les entreprises doivent recourir à des stratégies qualité qui répondent bien à leurs besoins et qui leur donnent la possibilité de proposer à leurs clients des produits et des services conformes aux spécifications. La maîtrise statistique de processus (MSP) est un outil qui permet aux entreprises de fabriquer des produits de plus grande qualité grâce à l'utilisation de méthodes statistiques d'analyse et de contrôle de l'évolution temporelle du processus.

Les cartes de contrôle ont joué un rôle très important dans cette démarche d'amélioration continue de la qualité des processus industriels. Depuis leur apparition dans les années '30 elles n'ont pas cessé d'évoluer mais en gardant un caractère univarié. Actuellement, suite à l'extraordinaire évolution de l'informatique et de l'électronique, on assiste à un changement majeur dans la manière dont les industriels abordent la maîtrise statistique de processus. Les cartes de contrôle multivariées, plus adéquates à décrire l'évolution réelle et souvent très complexe des processus industriels, sont de plus en plus utilisées. Elles ont l'avantage de permettre le contrôle simultané de plusieurs variables de processus et de prendre en compte la corrélation entre les paramètres du processus.

Nous avons, dans le chapitre I, fait une présentation sommaire des cartes de contrôle multivariées les plus utilisées. Un inconvénient majeur lié à l'utilisation de ces cartes de contrôle est l'absence de détails concernant la variable ou les variables responsable(s) de l'anomalie détectée durant le fonctionnement du processus. Ainsi, même si globalement on sait que le processus est hors contrôle, il est difficile de mettre en place les actions correctives nécessaires pour améliorer son fonctionnement. Plusieurs recherches ont été menées durant ces dernières années pour remédier à cet inconvénient et plusieurs solutions ont été proposées. Dans le chapitre II, nous avons regroupé quelques méthodes de détection parmi les plus représentatives pour leur principe de construction et leur contenu théorique. Nous avons effectué une présentation et une analyse comparative des ces méthodes et nous avons mis en évidence leurs avantages et leurs inconvénients. Par la suite, dans le chapitre III, nous avons proposé une nouvelle méthodologie de contrôle de processus en temps réel: la méthodologie FNAD (**F**iltrage **N**umérique et **A**nalyse **D**iscriminante) basée principalement sur l'utilisation des techniques de filtrage numérique de signal et sur l'analyse discriminante. Une fois mise en œuvre, cette méthode de contrôle de processus est très simple à exploiter et les résultats obtenus (la détection et le diagnostic de l'état de fonctionnement du processus) sont très facilement interprétables par les opérateurs. Un intérêt particulier a été accordé, dans le cadre de la méthodologie FNAD, à l'information disponible concernant le processus de fabrication (l'avis des experts et l'analyse historique des données disponibles dans une base de données utilisateur) et ceci constitue une différence majeure par rapport aux autres techniques de contrôle et de détection analysées dans le chapitre II. Il convient de mentionner ici que, la méthodologie de contrôle FNAD présente également l'avantage d'être en étroite liaison avec les démarches d'optimisation de processus. En effet, les informations stockées dans la base de données peuvent être issues des expérimentations visant l'optimisation du processus (comme la méthode des plans d'expériences par exemple). Durant ces essais plusieurs paramètres de processus sont modifiés et les résultats obtenus sont ensuite enregistrés pour des analyses ultérieures. On dispose donc d'un ensemble d'informations qui peut s'avérer très utile pour le diagnostic du processus car à travers ces essais la nature des relations entre les variables de processus et les variables qualitatives du produit est

identifiée. Nous considérons qu'il est très important de conserver toute cette information acquise à l'étape d'optimisation et d'en tirer profit pour le diagnostic et le suivi "on-line" du processus. Une attention spéciale a été accordée au filtrage numérique de signal. Nous avons effectué, d'abord, une analyse détaillée de deux causes assignables, souvent rencontrées en pratique: le saut en échelon et la dérive en rampe. Nous avons analysé ensuite les caractéristiques et les performances des différents filtres numériques afin de déterminer le type de filtre le plus adéquat pour la démarche de détection. L'efficacité de l'analyse discriminante en tant qu'outil d'aide à la prise de décision a été analysée dans le cas des variables indépendantes et corrélées. Les performances de la méthode FNAD ont été démontrées à l'aide d'un exemple pratique concernant un processus de fabrication du polyéthylène de faible densité dans des réacteurs à haute pression.

Deux nouvelles cartes de contrôle - la carte de contrôle par filtrage numérique (CCFN) et la carte de contrôle spectrale (CCS) – ont été proposées dans le chapitre IV. Le principe de construction de la carte CCFN repose sur l'utilisation des équations récurrentes définissant les filtres numériques RII (Réponse Impulsionnelle Infinie). Nous avons choisi l'utilisation des coefficients caractérisant les filtres passe-bas de type Butterworth et nous avons déterminé à l'aide de simulations les valeurs numériques pour les limites de contrôle de la carte dans le cas de différents filtres. Nous avons montré que les méthodes utilisées pour le calcul de la période opérationnelle moyenne (POM) des cartes EWMA et CUSUM ne peuvent pas être utilisées pour évaluer la POM de la carte CCFN. Des recherches devront encore être menées afin de trouver une expression analytique de calcul pour ces valeurs. Par des simulations, nous avons montré que les performances de la carte CCFN sont "légèrement" inférieures à celles de la carte EWMA dans le cas des dérèglages de très faibles amplitudes (inférieures ou égales à 1σ). Une carte de contrôle multivariée par filtrage numérique a été également proposée et les valeurs caractéristiques pour les limites de contrôle ont été déterminées par des simulations. Des méthodes de calcul analytique de la POM pour ce type de cartes manquent et devront faire l'objet de recherches. La construction de la carte de contrôle spectrale (CCS) est basée sur l'analogie temps-fréquence. Nous avons montré qu'il est possible d'utiliser l'information contenue dans les coefficients de la décomposition en série de Fourier d'un signal, afin de détecter les changements (sauts en échelon ou dérives en rampe) de la moyenne. A partir des considérations théoriques concernant la distribution des coefficients dans le spectre d'amplitude du signal nous avons déduit la formule analytique de la limite de contrôle sur la carte. Les performances de la carte de contrôle spectrale ont été calculées à partir de simulations et comparées avec celles de la carte EWMA. Nous avons constaté que la carte de contrôle spectrale présente de très bonnes performances dans la détection des dérèglages faibles, supérieures à celles de la carte EWMA notamment dans le cas des dérèglages de faibles amplitudes.

Pour conclure, nous considérons que la voie ouverte dans le domaine de la MSP par l'utilisation des techniques fréquemment utilisées dans d'autres domaines comme, le traitement numérique de signal, la spectrographie, l'imagerie médicale, etc... est très intéressante et mérite d'être explorée dans le cadre de futures recherches.

Annexes

- A.1 Evénements majeurs en Qualité
- A.2 Calcul des coefficients de régression biaisés
- A.3 Carte de contrôle pour des observations couplées
- A.4 Calcul des axes factoriels discriminants
- A.5 Formule récurrente pour un filtre ARMA
- A.6 Calcul de la POM_1 par la méthode des chaînes de Markov

A.1 Evénements majeurs dans le domaine de la Qualité

tableau 1 – Evénements majeurs dans le domaine de la Qualité

1875	Frederick W. Taylor énonce les principes du "Management Scientifique" qui consiste à diviser un travail complexe en tâches plus simples et plus faciles à réaliser. Gilbreth et Gantt proposent ultérieurement des contributions similaires. L'objectif principal reste la productivité.
1900-1930	Henry Ford raffine d'avantage les méthodes d'amélioration de la productivité et de la qualité. Ford propose des concepts comme l'autotest, l'inspection en ligne, etc.
1901	Création du Laboratoire National d'Essais (LNE) en France pour répondre aux besoins de mesures et d'essais de l'industrie.
1906	Commencement de la normalisation internationale, avec la création de la Commission Electrotechnique Internationale (CEI).
1907-1908	AT&T pratique l'inspection et le test des produits et des matières premières systématiquement.
1908	W.S.Gosset (écrivain sous le pseudonyme "Student") propose la distribution t (de Student) - suite à ses travaux en contrôle qualité chez Guinness Brewery.
1915-1919	Le gouvernement Britannique commence un programme de certification pour les fournisseurs.
1918	Création de la Commission Permanente de Standardisation en France, pour "étudier toutes les mesures susceptibles d'assurer l'unification des types dans la construction mécanique et métallique". Elle disparaîtra lors de la création de l'AFNOR en 1926.
1919	"Technical Inspection Association" est constituée en Angleterre; ultérieurement elle portera le nom d' "Institute of Quality Assurance".
1920	AT&T Bell Laboratoires jettent les bases d'un département Qualité qui envisage l'amélioration de la qualité, l'inspection, le test et la fiabilité des produits.

	B.P.Dudding et General Electrics en Angleterre utilisent des méthodes statistiques afin de contrôler la qualité des lampes électriques.
	Fisher élabore l'analyse de la variance et réalise un plan d'expérience statistique applicable à l'étude de la culture des pommes de terre et des céréales.
1922	Radford publie "The Control of Quality in Manufacturing" où la qualité est présentée comme indépendante du management.
1922-1923	R.A.Fisher publie une série d'articles concernant les plans d'expériences et leurs applications dans l'agriculture.
1924	Création du Département Qualité Bell Telephone Laboratories avec Shewhart, Dodge, Roming puis Deming, Juran et Edwards qui deviendra le "père" de L'ASSURANCE QUALITE.
	W.A.Shewhart propose la notion de carte de contrôle dans un mémorandum technique à Bell Laboratoires.
1926	Création de l'Association Française de Normalisation (AFNOR). Création de la fédération internationale des associations nationales de normalisation (ISA) qui cessera ses activités en 1942 (en raison de la Seconde Guerre mondiale).
1928	Les plans d'acceptation par échantillonnage sont proposés par H.F. Dodge et H.G. Roming (Bell Laboratoires). Juran présente un cours de formation professionnelle nommé Quality control.
1931	Shewhart publie le "Contrôle Economique des Produits Manufacturés" qui permet une approche scientifique de la qualité, mettant l'accent sur l'utilité des méthodes statistiques dans la production industrielle et l'avantage de l'utilisation des cartes de contrôle.
1932	W.A.Shewhart donne des cours sur les méthodes statistiques de contrôle et les cartes de contrôle à l'Université de Londres.
1932-1933	L'industrie textile et du bois britannique et l'industrie chimique allemande commencent à s'intéresser aux plans d'expériences pour l'optimisation de leurs procédés de fabrication.
1933	Royal Statistical Society jette les bases de la section de recherche industrielle et agricole (Industrial and Agricultural Research Section). Premier Salon de la Qualité Française à Paris.
1934	Développement du diagramme de Pareto pour la classification des défauts selon leur gravité par Talacko et Veslo.
1938	W.E.Deming invite Shewhart pour presenter des seminaires à l'U.S. Department of Agriculture Graduate School.
1940	Publication des Military Standards par le Département de la Défense des Etats-Unis.
1940-1943	Bell Labs élaborent les premières normes de plans d'échantillonnage pour l'armée américaine.
1942	En Angleterre est fondé le "Ministère de Conseils pour des Méthodes Statistiques et Contrôle de la Qualité". Programme de formation à la Maîtrise de la Qualité par Deming et Juran dans les usines d'armement US.
1944	Les premières publications "Industrial Quality Control" apparaissent.
1945	Feigenbaum publie son article "Quality as a Management".
1946	Création de l'American Society for Quality Control (ASQC). Deming est invité en Japon pour aider les forces d'occupation à

	reconstruire l'industrie nippone.
1946-1949	L'Union des Scientifiques et des Ingénieurs Japonais (JUSE) est créé. Deming est invité en Japon pour des séminaires au sujet du contrôle statistique de processus.
1947	Création de l'ISO, fédération mondiale d'organismes nationaux de normalisation ayant "pour mission de favoriser le développement de la normalisation et des activités connexes dans le monde, en vue de faciliter les échanges de biens et de services entre les nations et de développer la coopération dans les domaines intellectuel, scientifique, technique et économique". L'ISO entre officiellement en fonction le 23 février 1947.
1948	Le professeur G. Taguchi commence des études sur des plans d'expériences.
1950	Deming forme des managers dans le domaine industriel au Japon; les méthodes de contrôle statistique de processus commencent à être largement enseignées en Japon. Le professeur K. Ishikawa élabore le diagramme en arêtes de poisson. Des textes de référence dans le domaine du contrôle statistique sont publiés par Eugene Grant et A.J.Duncan
1951	Lancement du concept Total Quality Control (TQC) par Feigenbaum. Parution du "Quality Control Handbook" de Juran. Le "Prix Deming" est attribué pour des contributions importantes dans le domaine de la qualité et du contrôle de la qualité Mise au point de l'AMDEC (Analyse des Modes de Défaillance et de leur Criticité). Publication de la première norme ISO sous le titre "Température normale de référence des mesures industrielles de longueur "
1951+	G.E.P.Box et K.B.Wilson publient des œuvres de référence concernant l'utilisation des plans d'expériences et de la méthode de surface de réponse pour l'optimisation de processus; les applications dans le domaine de la chimie se développent ultérieurement.
1954	Le docteur Joseph M. Juran est invité en Japon pour des séminaires au sujet du management de la qualité. Le statisticien anglais E.S.Page élabore la carte de contrôle CUSUM (Cumulative SUM).
1955	Diffusion par Ishikawa de la carte de contrôle (créée par Shewart en 1924).
1957	J.M.Juran et F.M.Gryna publient "Quality Control Handbook". Création de l'Association Française pour le Contrôle Industriel et la Qualité (AFCIQ).
1959	La revue "Technometrics" est apparue avec J. Stuart Hunter comme éditeur fondateur. S. Roberts élabore la carte de contrôle EWMA (Exponentially Weighted Moving Average). Publication de la première norme d'assurance de la qualité par l'armée américaine : la norme MIL-Q-9858.
1960	G.E.P.Box et J.S.Hunter écrivent des articles de référence sur les plans factoriels 2^{k-p} . Le concept de Cercle de Contrôle de la Qualité est introduit en Japon par K.Ishikawa.
1961	Le Conseil National de la Qualité et de la Productivité est formé en

	<p>Angleterre.</p> <p>Des cours de contrôle statistique de processus sont fréquemment enseignés dans le cadre des Programmes Académiques d'Ingénierie Industrielle.</p> <p>Le concept de Zéro Défauts est introduit aux Etats Unis par Crosby dans le cadre des programmes spatiaux APOLLO.</p> <p>Création de l'Association Française des Qualiticiens (AFQ).</p>
1969	<p>La revue "Industrial Quality Control" cesse de paraître et elle est remplacée par "Quality Progress" et "Journal of Quality Technology".</p>
1970	<p>Les Etats-Unis imposent par la loi l'obligation de respecter des critères d'AQ pour la construction des centrales nucléaires.</p> <p>L'AFNOR lance l'étude des normes relatives à la gestion de l'Assurance Qualité.</p>
1971	<p>Publication du "Manuel des Cercles de Qualité" par la JUSE. Les cercles de qualité sont à la fois des groupes de réflexion et des groupes d'auto-formation.</p>
1974	<p>EDF impose la mise en place d'une organisation de la qualité pour la construction des centrales nucléaires (160 fournisseurs principaux).</p> <p>Création des premiers cercles de qualité américains, à partir des ouvrages d'Ishikawa.</p>
1975	<p>Création du Service de la Qualité des Produits Industriels et de la Normalisation (SQUALPI), actuellement sous la direction de la Qualité pour l'Industrie et de la Normalisation au sein du ministère chargé de l'Industrie.</p>
1975-1978	<p>Des livres de plans d'expériences orientés vers les ingénieurs et les scientifiques commencent être publiés.</p> <p>L'intérêt pour les Cercles de Qualité augmente en Amérique du Nord. Cela conduit au mouvement pour le Management de la Qualité Totale.</p>
1977	<p>Le français Michel Vigier reçoit le prix de la Division Automobile de l'ASQC (seul non américain à avoir reçu ce prix).</p>
1978	<p>Création de l'Association Française pour l'Analyse de Valeur (AFAV): elle a pour mission de promouvoir l'utilisation de l'analyse de la valeur.</p> <p>Premiers cercles de qualité aux usines Citroën de Rennes.</p>
1979	<p>Lancement de l'étude de normes internationales d'assurance de la qualité par l'ISO (une trentaine de pays y participe).</p> <p>Création au sein de l'ISO du Technical Comity 176 (TC176). Il a en charge la normalisation dans le domaine de la qualité.</p> <p>Création du Réseau National d'Essais (RNE) en France.</p>
1980	<p>Les plans d'expériences sont adoptés par un grand nombre de branches industrielles dont l'électronique, l'aérospatiale, l'industrie de semi-conducteurs et l'industrie de l'automobile.</p> <p>Les travaux de G. Taguchi au sujet des plans d'expériences sont publiés aux Etats Unis pour la première fois.</p> <p>Publication par l'AFNOR de la norme NF X 50-110 (norme sur la gestion de la qualité). Elle est annulée en 1987 au profit de la norme NF EN 29.004 (ISO 9004).</p>
1981	<p>Présentation de l'action de Deming sur la chaîne de télévision NBC.</p> <p>Création de l'Association Française pour les Cercles de Qualité (AFCERQ)</p> <p>Intervention télévisée de Juran sur CBS "Alors, si le Japon le peut,</p>

	pourquoi pas nous?"
1982	Publication de la norme BS 5750, d'assurance de la qualité en Angleterre.
1983	Publication par l'AFNOR du recueil des normes françaises : "Gérer et assurer la qualité".
1986	Box et autres visitent le Japon et remarquent l'utilisation extensive des plans d'expériences et des méthodes statistiques par les japonais.
1987	Naissance de la série des normes ISO 9000. En 1994 elles subissent une première révision et une seconde en 2000.
	Création du prix national américain de la qualité sous l'impulsion de Deming "Malcolm BALDRIGE National Quality Award".
1988	Création de l'Association Française d'Assurance de la Qualité (AFAQ). Création de l'European Foundation for Quality Management (EFQM).
1989	La revue "Quality Engineering" apparaît.
1990	L'intérêt pour la certification ISO 9000 augmente aux Etats Unis; les candidats pour le prix Baldrige augmentent lentement; plusieurs pays sponsorisent des prix pour la Qualité basés sur des critères de type Baldrige.
1991	Le Mouvement Français pour la Qualité (MFQ) est créé à l'initiative du MEDEF et du Ministère chargé de l'Industrie. Il naît de la fusion de l'AFCIQ, de l'AFQ et de l'AFCERQ. L'EFQM (en collaboration avec l'European Organization for Quality, et la Commission Européenne) crée l'European Quality Award, pour inciter les entreprises européennes à jouer un rôle actif en terme de qualité face aux très puissants Américains et Japonais.
1992	Le Ministère de l'industrie et le MFQ créent le Prix Français de la Qualité pour récompenser les efforts et les résultats obtenus par les entreprises dans le cadre de la démarche Qualité.
1993	Création du RUFEREQ (le Réseau Universitaire Français d'Enseignement et de la Recherche de la Qualité).
1994	Création du Comité Français d'Accréditation (COFRAC). Le COFRAC regroupant Le RNE et le Bureau National de Métrologie. Evolution des normes ISO 9000. Le premier congrès Qualité et Sécurité de Fonctionnement organisé par RUFEREQ à l'UTC.
1995	Loi 95-96 du 01/02/95 concernant le Marquage CE. Lancement du Mois de la Qualité par le MFQ.
1996	Publication des premières normes de la famille ISO 14000 pour la certification du système de management environnemental.
1997	Le deuxième congrès Qualité et Sécurité de Fonctionnement à l'ISTIA Angers.
1999	Le troisième congrès Qualité et Sécurité de Fonctionnement à l'ENSAM Paris

2000	Le passage des normes ISO 9000 vers la version 2000
2001	Le quatrième congrès Qualité et Sûreté de Fonctionnement à Annecy.
2003	Le cinquième congrès Qualité et Sûreté de Fonctionnement aura lieu à Nancy.

A.2 Calcul des coefficients de régression biaisés

Houshmand et Javahery [54] analysent la situation où les variables sont fortement corrélées, ce qui se traduit par l'impossibilité du calcul de l'inverse de la matrice de variance-covariance et rend difficile la détermination des coefficients de régression stables. A la suite, l'effet des variables indépendantes ne peut plus être déterminé avec précision et cette situation est connue sous le nom d'effet de colinéarité multiple.

Les auteurs proposent l'utilisation d'une modification de la méthode des moindres carrés qui donne des coefficients de régression biaisés, mais qui ont l'avantage d'avoir une variabilité plus réduite que les estimateurs basés sur la méthode des moindres carrés. Par conséquent, les modèles ainsi construits sont plus stables. Pour atteindre cet objectif, ils ont introduit une constante $C > 0$ dans l'équation des moindres carrés:

$$(X^T X + CI)\hat{\beta}_R(C) = X^T y \quad (A.2.1)$$

où y est le vecteur ($n \times 1$) de la variable dépendante, X est la matrice ($n \times p$) des variables indépendantes et $\hat{\beta}_R(C)$ est le vecteur ($p \times 1$) des coefficients de régression.

Afin de mieux contrôler les erreurs d'arrondissement et pour comparer les coefficients de régression sur une même échelle, les données sont centrées et réduites:

$$x_{ij}^0 = \frac{x_{ij} - \bar{x}_j}{\sqrt{s_{jj}}}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, p \quad (A.2.2)$$

$$y_i^0 = \frac{y_i - \bar{y}}{\sqrt{s_{yy}}}, \quad i = 1, 2, \dots, n \quad (A.2.3)$$

où x_{ij}^0 et y_i^0 sont des variables centrées et réduites, et s_{jj} et s_{yy} sont définies ainsi:

$$s_{jj} = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \quad (A.2.4)$$

$$s_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (A.2.5)$$

Si l'on utilise les transformations précédentes, les coefficients de régression biaisés sont obtenus à l'aide de l'équation:

$$(r_{xx} + CI)b^R = r_{yx} \quad (A.2.6)$$

où r_{xx} représente la matrice de corrélation pour les variables indépendantes X , r_{yx} le vecteur de corrélation entre la variable dépendante Y et les variables indépendantes X et b^R le vecteur des coefficients de régression standardisés.

Si la constante C est nulle alors la régression avec les coefficients biaisés se transforme dans la régression linéaire multiple classique qui repose sur la méthode des moindres carrés. Plus la constante C est grande plus le biais dans les coefficients de régression est augmenté et donc plus ils seront loin de la valeur réelle. Il est essentiel que C reste la plus petite possible mais en même temps assez importante pour avoir une bonne estimation.

La constante C peut être choisie soit pour maximiser la stabilité du modèle soit pour maximiser le pouvoir prédictif du modèle. Dans ce cas elle est déterminée afin de minimiser la fonction somme des erreurs de prédiction au carré (SEPC) dont l'expression est:

$$SEPC = \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{1 - h_{ii}} \right)^2$$

où $y_i - \hat{y}_i$ est le résidu de la régression calculé par la méthode des moindres carrés après avoir éliminé l'observation i , h_{ii} est le i -ème élément dans la diagonale de la matrice de projection (ou matrice chapeau): $H = X(X^T X)^{-1} X^T$.

Si on généralise la fonction SEPC pour le cas de la régression biaisée on obtient l'expression suivante:

$$SEPC_R(C) = \sum_{i=1}^n \left(\frac{y_i - x_i^T \hat{\beta}_R(C)}{1 - x_i^T (X^T X + CI)^{-1} x_i} \right)^2 \quad (A.2.7)$$

Malheureusement cette fonction n'est pas strictement convexe, donc on peut trouver un minimum local unique. Montgomery et Friedman [86] ont suggéré de rechercher le coefficient C dans l'intervalle $0 \leq C \leq 2$ et de choisir la valeur de C qui donne un minimum local dans cet intervalle.

Houshmand et Javahery affirment que les résidus obtenus après la régression de chaque variable dans toutes les autres par cette méthode peuvent être représentés sur une carte classique de contrôle (Shewhart ou EWMA) pour le diagnostic de la variable responsable du dérèglement. Ils ont démontré aussi à partir des simulations que la POM (Période Opérationnelle Moyenne) de la carte ainsi obtenue est meilleure que la POM de la carte T^2 Hotelling. L'inconvénient majeur de la méthode, à part l'incertitude dans le calcul du coefficient C optimal, est le grand nombre de cartes de contrôle qui doivent être surveillées quand le nombre de variables devient important (une carte de contrôle par variable).

A.3 Carte de contrôle pour des observations couplées

Cette carte de contrôle a été proposée par Tracy, Young et Mason [124] et elle est applicable uniquement pour des observations bivariées couplées. C'est le cas, en particulier, des mesures effectuées sur un processus chimique, où les résultats obtenus dans un certain laboratoire sont vérifiés par un laboratoire principal. Si les résultats

obtenus dans les deux laboratoires sont en concordance, alors le meilleur estimateur de la moyenne du processus est la moyenne de deux mesures. La moyenne peut être utilisée pour le suivi du processus à l'aide d'une carte de contrôle de type Shewhart. Le problème est qu'une telle procédure de suivi nécessite des observations indépendantes et cette condition n'est pas respectée car les mesures sont effectuées sur les mêmes échantillons dans les deux cas. Quand les résultats des deux laboratoires sont en concordance, l'observation de la variabilité du produit est l'objectif principal. Quand, par contre, les résultats obtenus par les deux laboratoires ne concordent pas, la conclusion est qu'il s'est passé quelque chose pendant les mesures effectuées dans un des laboratoires ou dans les deux. Donc, les résultats d'un laboratoire au moins sont biaisés et il faut faire attention à leur utilisation directe pour le suivi de la moyenne du processus.

Considérons le cas où un échantillon est analysé par un laboratoire et ensuite il est envoyé dans un laboratoire principal pour la vérification des résultats. Deux mesures sont effectuées sur le même échantillon. Théoriquement les deux observations couplées doivent être en concordance, mais en pratique ce n'est pas toujours le cas. Les mesures peuvent être différentes à cause de l'environnement, des instruments utilisés pour effectuer les mesures, des opérateurs, etc....

Si l'on note les n observations par X_{ij} , $i=1,2,\dots,n$ et $j = 1$ pour le premier laboratoire et $j = 2$ pour le deuxième, les résultats des mesures peuvent être représentés par le modèle suivant:

$$X_{ij} = \mu + d_i + e_{ij} \quad i = 1,2,\dots,n; \quad j = 1,2; \quad (\text{A.3.1})$$

où μ est la moyenne globale du processus mesuré, d_i est la déviation aléatoire de l'échantillon i (avec moyenne nulle) et e_{ij} est l'erreur aléatoire de mesure (avec moyenne nulle).

La composante aléatoire d_i est considérée non-corrélée avec e_{ij} et a sa variance σ_d^2 donne la variance du processus. D'une manière analogue, la variance de la composante e_{ij} , notée σ_{ej}^2 donne la variance des mesures du laboratoire.

Le modèle présenté dans l'équation (A.3.1) est le modèle utilisé par Grubbs [40, 41] pour estimer la reproductibilité de la mesure.

Si chaque couple d'observations dans l'équation A.3.1 est représenté par un vecteur bidimensionnel $X = (X_1, X_2)^T$, on peut considérer que ce vecteur est décrit par une fonction de densité de probabilité normale bidimensionnelle:

$$f(X_1, X_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{T^2}{2}\right)$$

des paramètres:

$$\mu = (\mu_1, \mu_2)^T \text{ et } \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} :$$

où $\sigma_{12} = \sigma_{21}$ représentent la covariance entre X_1 et X_2 , et σ_1^2 et σ_2^2 représentent les variances de X_1 et de X_2 respectivement et où:

$$T^2 = \frac{1}{1-\rho^2} \times \left[\frac{(X_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(X_1 - \mu_1)(X_2 - \mu_2)}{\sigma_1 \sigma_2} + \frac{(X_2 - \mu_2)^2}{\sigma_2^2} \right] \quad (\text{A.3.2})$$

a une distribution de khi-deux avec deux degrés de liberté. Dans l'équation (A.3.2) ρ représente le coefficient de corrélation entre les deux variables X_1 et X_2 .

Si l'on considère que d_i et e_{ij} ont des moyennes nulles, alors μ_1 et μ_2 sont égales à μ . Ces variances sont la somme de deux composantes: la variation du processus σ_d^2 et la variation de l'erreur de mesure du laboratoire $\sigma_{e_j}^2$. Ainsi:

$$\begin{cases} \sigma_1^2 = \sigma_d^2 + \sigma_{e1}^2 \\ \sigma_2^2 = \sigma_d^2 + \sigma_{e2}^2 \end{cases}$$

Dans chacune de ces formules il est estimé que la variation du processus est dominante.

Considérons le cas où les deux laboratoires disposent de conditions excellentes de travail ($\sigma_{e1}^2 = \sigma_{e2}^2 = \sigma_e^2$). La différence $X_1 - X_2$, est une mesure de la concordance entre les deux laboratoires, et on peut estimer que la distribution de cette différence est centrée sur zéro. Considérons aussi que les variances de X_1 et de X_2 sont identiques, et donc:

$$\sigma_1^2 = \sigma_2^2 = \sigma^2 = \sigma_d^2 + \sigma_e^2$$

Il en résulte que:

$$\text{Var}(X_1 - X_2) = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2 = 2\sigma^2 - 2\rho\sigma^2 = 2\sigma^2(1 - \rho) \quad (\text{A.3.3})$$

où $\rho = \frac{\sigma_{12}}{\sigma_1\sigma_2}$ est le coefficient de corrélation entre X_1 et X_2 .

Par conséquent, quand les mesures des deux laboratoires sont en parfaite concordance ($\rho = 1$) il n'y a pas de variation pour la différence $X_1 - X_2$.

A partir des données fournies par les deux laboratoires, le meilleur estimateur de la moyenne μ est $\bar{X} = \frac{1}{2}(X_1 + X_2)$. La variance de l'estimateur de la moyenne est:

$$\text{Var}(\bar{X}) = \frac{1}{4}\sigma_1^2 + \frac{1}{4}\sigma_2^2 + 2 \cdot \frac{1}{2} \cdot \frac{1}{2} \rho\sigma_1\sigma_2 = \frac{1}{2}\sigma^2 + \frac{1}{2}\rho\sigma^2 = \frac{1}{2}\sigma^2(1 + \rho) \quad (\text{A.3.4})$$

Si $\rho = 1$, la variance de la moyenne est $\sigma^2 = \sigma_d^2 + \sigma_e^2$ et elle est dominée par le variance du processus.

Les équations (A.3.3) et (A.3.4) nous suggèrent une relation intéressante entre le coefficient de corrélation, la concordance des laboratoires et la variation du processus.

Quand les deux laboratoires sont en concordance et la variation du processus est faible, on ne doit pratiquement pas avoir de corrélation entre les mesures (car les observations couplées seront groupées dans un intervalle étroit et une tendance linéaire ne sera pas visible). Si les laboratoires sont en concordance et la variation du processus est grande, une forte corrélation entre les mesures sera présente. Finalement, quand les deux laboratoires ne sont pas en concordance, la corrélation entre les mesures sera faible pour toutes les valeurs de la variation du processus.

Si dans l'équation (A.3.2) on fait $T^2 = c$, $c > 0$, on obtient dans le plan formé par X_1 et X_2 une ellipse centrée sur $\mu = (\mu_1, \mu_2)^T$. Après avoir centré et réduit les variables X_1 et X_2 :

$$Y_1 = \frac{X_1 - \mu_1}{\sigma_1} \quad \text{et} \quad Y_2 = \frac{X_2 - \mu_2}{\sigma_2}$$

l'équation de l'ellipse dévient:

$$T^2 = \frac{1}{1 - \rho^2} (Y_1^2 - 2\rho Y_1 Y_2 + Y_2^2) = c \quad (\text{A.3.5})$$

Si l'on fait une rotation de 45° dans l'équation (A.3.5) en utilisant la transformation orthogonale:

$$t_1 = \frac{Y_1 + Y_2}{\sqrt{2}} \quad \text{et} \quad t_2 = \frac{Y_2 - Y_1}{\sqrt{2}} \quad (\text{A.3.6})$$

on obtient:

$$T^2 = \frac{t_1^2}{1 + \rho} + \frac{t_2^2}{1 - \rho} = c \quad \text{ou} \quad T^2 = \frac{t_1^2}{c(1 + \rho)} + \frac{t_2^2}{c(1 - \rho)} = 1 \quad (\text{A.3.7})$$

L'équation (A.3.7) représente l'équation d'une ellipse centrée dans l'origine du système des coordonnées et ayant l'axe principal de longueur $2\sqrt{c(1 + \rho)}$ et l'axe secondaire de longueur $2\sqrt{c(1 - \rho)}$.

Il est possible de démontrer que t_1 et t_2 représentent les composantes principales de l'ensemble des observations dans le plan formé par X_1 et X_2 . Par conséquent, s'il y a une concordance entre les laboratoires, t_1 peut être utilisée pour le suivi de la moyenne du processus. La deuxième composante principale t_2 peut être utilisée pour vérifier la concordance des résultats entre les deux laboratoires. Quand $\rho = 1$, la région de contrôle elliptique se réduit à un cercle, où la variation au long de l'axe t_1 est due à σ_d^2 et la variation au long de l'axe t_2 est due à σ_e^2 . Quand ρ s'approche de 1, la longueur de l'axe t_1 augmente par rapport à la longueur de l'axe t_2 ce qui nous indique la dominance de la variation du processus.

En résumé, un déplacement au long de l'axe t_2 représente une non-concordance entre les deux laboratoires, alors qu'un déplacement au long de l'axe t_1 indique la modification de la moyenne du processus.

A.4 Calcul des axes factoriels discriminants

Le problème formulé dans l'équation (3.12) est un problème classique de maximum du rapport de deux formes quadratiques. Nous présentons la solution qui utilise les multiplicateurs de Lagrange.

Il faut remarquer d'abord que \mathbf{u} est défini à une constante multiplicative près. On peut fixer la contrainte $\mathbf{uTu}^T = \text{cte}$ et il faut maximiser \mathbf{uBu}^T avec $\mathbf{uTu}^T = \text{cte}$.

Nous introduisons le multiplicateur λ , et il devient:

$$\frac{\partial}{\partial \mathbf{u}} [\mathbf{uBu}^T - \lambda(\mathbf{uTu}^T - \text{cte})] = 0 \leftrightarrow 2\mathbf{Bu}^T = 2\lambda\mathbf{Tu}^T \quad (\text{A.4.1})$$

et, si T est inversible:

$$\mathbf{T}^{-1}\mathbf{Bu}^T = \lambda\mathbf{u}^T \quad (\text{A.4.2})$$

\mathbf{u}^T est, donc, vecteur propre de $\mathbf{T}^{-1}\mathbf{B}$.

Si l'on multiplie les deux membres de l'équation (A.4.1) par \mathbf{u} à gauche, il devient:

$$\mathbf{uBu}^T = \lambda\mathbf{uTu}^T, \text{ soit } \lambda = \frac{\mathbf{uBu}^T}{\mathbf{uTu}^T} \text{ qui est justement la quantité que l'on veut maximiser.}$$

Le premier axe factoriel discriminant \mathbf{u}_1 est donc le vecteur propre de $\mathbf{T}^{-1}\mathbf{B}$ correspondant à la plus grande valeur propre λ_1 .

A.5 Formule récurrente y en fonction de x pour un filtre ARMA

Pour déduire la formule générale de l'équation récurrente y en fonction de x pour un filtre ARMA d'ordre L nous considérons au début la cas où l'ordre du filtre est L = 2. Dans ce cas l'équation (4.7) devient:

$$\mathbf{b}_0\mathbf{y}_k = \sum_{i=0}^2 \mathbf{a}_i\mathbf{x}_{k-i} + \sum_{j=1}^2 \mathbf{b}_j\mathbf{y}_{k-j} \quad (\text{A.5.1})$$

et si on développe cette équation en fonction de \mathbf{k} , en considérant $\mathbf{b}_0 = 1$, on peut écrire:

$$\begin{aligned} \mathbf{y}_0 &= \mathbf{a}_0\mathbf{x}_0 = \mathbf{c}_0\mathbf{x}_0 \\ \mathbf{y}_1 &= \mathbf{a}_0\mathbf{x}_1 + \mathbf{a}_1\mathbf{x}_0 + \mathbf{b}_1\mathbf{y}_0 = \mathbf{a}_0\mathbf{x}_1 + \mathbf{a}_1\mathbf{x}_0 + \mathbf{b}_1\mathbf{a}_0\mathbf{x}_0 = \mathbf{a}_0\mathbf{x}_1 + (\mathbf{a}_1 + \mathbf{b}_1\mathbf{a}_0)\mathbf{x}_0 = \mathbf{c}_0\mathbf{x}_1 + \mathbf{c}_1\mathbf{x}_0 \\ &\quad \text{où : } \mathbf{c}_1 = \mathbf{a}_1 + \mathbf{b}_1\mathbf{a}_0 = \mathbf{a}_1 + \mathbf{b}_1\mathbf{c}_0 \\ \mathbf{y}_2 &= \mathbf{a}_0\mathbf{x}_2 + \mathbf{a}_1\mathbf{x}_1 + \mathbf{a}_2\mathbf{x}_0 + \mathbf{b}_1\mathbf{y}_1 + \mathbf{b}_2\mathbf{y}_0 = \mathbf{a}_0\mathbf{x}_2 + \mathbf{a}_1\mathbf{x}_1 + \mathbf{a}_2\mathbf{x}_0 + \mathbf{b}_1(\mathbf{c}_0\mathbf{x}_1 + \mathbf{c}_1\mathbf{x}_0) + \mathbf{b}_2\mathbf{c}_0\mathbf{x}_0 = \\ &= \mathbf{a}_0\mathbf{x}_2 + (\mathbf{a}_1 + \mathbf{b}_1\mathbf{c}_0)\mathbf{x}_1 + (\mathbf{a}_2 + \mathbf{b}_1\mathbf{c}_1 + \mathbf{b}_2\mathbf{c}_0)\mathbf{x}_0 = \mathbf{c}_0\mathbf{x}_2 + \mathbf{c}_1\mathbf{x}_1 + \mathbf{c}_2\mathbf{x}_0 \end{aligned}$$

$$\begin{aligned}
 & \text{où : } c_2 = a_2 + b_1c_1 + b_2c_0 \\
 y_3 &= a_0x_3 + a_1x_2 + a_2x_1 + b_1y_2 + b_2y_1 = a_0x_3 + a_1x_2 + a_2x_1 + b_1(c_0x_2 + c_1x_1 + c_2x_0) + \\
 & + b_2(c_0x_1 + c_1x_0) = c_0x_3 + (a_1 + b_1c_0)x_2 + (a_2 + b_1c_1 + b_2c_0)x_1 + (b_1c_2 + b_2c_1)x_0 = \\
 & = c_0x_3 + c_1x_2 + c_2x_1 + c_3x_0 \\
 & \text{où : } c_3 = b_1c_2 + b_2c_1 \\
 y_4 &= a_0x_4 + a_1x_3 + a_2x_2 + b_1y_3 + b_2y_2 = a_0x_4 + a_1x_3 + a_2x_2 + \\
 & + b_1(c_0x_3 + c_1x_2 + c_2x_1 + c_3x_0) + b_2(c_0x_2 + c_1x_1 + c_2x_0) = a_0x_4 + (a_1 + b_1c_0)x_3 + \\
 & + (a_2 + b_1c_1 + b_2c_0)x_2 + (b_1c_2 + b_2c_1)x_1 + (b_1c_3 + b_2c_2)x_0 = \\
 & = c_0x_4 + c_1x_3 + c_2x_2 + c_3x_1 + c_4x_0 \\
 & \text{où : } c_4 = b_1c_3 + b_2c_2
 \end{aligned}$$

En généralisant les formules précédentes l'on déduit que :

$$y_k = \sum_{i=0}^k c_i x_{k-i}, \quad \text{où } c_i = \begin{cases} a_i + \sum_{j=1}^i b_j c_{i-j} & , i \leq 2 \\ \sum_{j=1}^L b_j c_{i-j} & , i > 2 \end{cases} \quad (\text{A.5.2})$$

Selon le même raisonnement, on peut déduire facilement les formules $y = f(x)$ pour tout ordre L du filtre, c'est-à-dire :

$$y_k = \sum_{i=0}^k c_i x_{k-i}, \quad \text{où } c_i = \begin{cases} a_i + \sum_{j=1}^i b_j c_{i-j} & , i \leq L \\ \sum_{j=1}^L b_j c_{i-j} & , i > L \end{cases} \quad (\text{A.5.3})$$

A.6 Calcul de la POMI par la méthode des chaînes de Markov

On divise l'intervalle compris entre les deux limites de contrôle de la carte dans un nombre $l = 2m + 1$ (où $m=1,2,\dots$) de sous-intervalles de longueur 2δ chacun. On note S_i le centre du sous-intervalle i (figure A.1). On dit que y_k se trouve à l'instant k dans l'état de transition i si:

$$S_i - \delta < y_k \leq S_i + \delta \quad (\text{A.6.1})$$

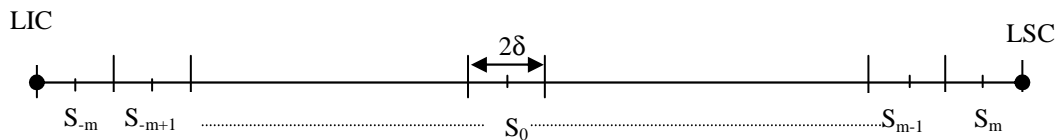


figure A.1

On considère que y_k se trouve dans l'état absorbant a si y_k dépasse une des deux limites de contrôle de la carte ($y_k < LIC$ ou $y_k > LSC$).

Le processus est considéré sous-contrôle chaque fois que y_k se trouve dans un état de transition. Le processus est hors-contrôle si y_k se trouve dans l'état absorbant a .

La distribution de la POM de la carte de contrôle EWMA est complètement déterminée si l'on connaît le vecteur de probabilités initiales et la matrice de transitions. Le vecteur des probabilités initiales est:

$$p_{init}^T = (p_{-m}, \dots, p_{-1}, p_0, p_1, \dots, p_m | 0) = (p^T | 0) \quad (A.6.2)$$

où p_j représente la probabilité que y soit dans l'état de transition j au départ. Il faut remarquer que la probabilité p_a est nulle parce que la statistique y est considérée sous-contrôle au départ.

En pratique, le vecteur des probabilités initiales contient soit un seul élément égal à 1 (représentant l'état de démarrage) et les autres nuls, soit il sera un vecteur de probabilités de type « en-cours » c'est à dire un vecteur où la probabilité de chaque état de transition est non-nulle.

La matrice de transitions, peut être divisée en plusieurs sous-matrices et est représentée ci-dessus :

$$P = \begin{pmatrix} R & (I - R)\bar{1} \\ \bar{0}^T & 1 \end{pmatrix} \quad (A.6.3)$$

où la sous-matrice R contient les probabilités de transitions d'un état de transition dans un autre, I est la matrice unitaire, $\bar{1}$ est un vecteur colonne dont tous les éléments sont égaux à 1 et $\bar{0}^T$ est un vecteur ligne dont tous les éléments sont nuls.

Si y_k est dans l'état i , alors on considère que $y_k = S_i$. Cette approximation nous donne la possibilité d'écrire la probabilité de transition d'un état i dans un état j à l'aide de la formule ci-dessous:

$$p_{ij} = P(y_k \in \text{Intervalle } j / y_{k-1} \in \text{Intervalle } i) = \\ = P \left[\frac{S_j - \delta - (1-r)S_i}{r} < x_k \leq \frac{S_j + \delta - (1-r)S_i}{r} \right], \quad j = -m, -m+1, \dots, m \quad (A.6.4)$$

Comme la somme des éléments situés sur les lignes de la matrice de transitions doit être égale à 1, les probabilités de transition d'un état sous-contrôle dans un état hors-contrôle sont trouvées par soustraction.

Pour le cas particulier où les observations x sont normalement distribuées $x \sim N(\mu_x, \sigma_x)$, les probabilités de transitions entre les états sous-contrôle sont calculées à l'aide de la formule:

$$p_{ij} = \Phi \left[\frac{S_j + \delta - (1-r)S_i - r\mu_x}{r\sigma_x} \right] - \Phi \left[\frac{S_j - \delta - (1-r)S_i - r\mu_x}{r\sigma_x} \right] \quad (\text{A.6.5})$$

où Φ représente la fonction de répartition normale centrée et réduite.

La matrice de transitions à l'étape t est très utile pour évaluer la distribution de la période opérationnelle (PO), car elle contient les probabilités de transition d'un état à l'autre en t pas ou étapes:

$$P^{(t)} = \begin{pmatrix} R^t & (I - R^t)\vec{1} \\ \vec{0}^T & 1 \end{pmatrix} \quad (\text{A.6.6})$$

Par la suite: $P(\text{PO} \leq t) = p^T (I - R^t) \vec{1}$

et

$$P(\text{PO} = t) = p^T (R^{t-1} - R^t) \vec{1} \quad (\text{A.6.7})$$

Par conséquent, utilisant l'équation (A.6.7) la POM basée sur $l=2m+1$ états est calculée à l'aide de la formule:

$$\begin{aligned} \text{POM}(l) &= \sum_{t=1}^{\infty} tP(\text{PO} = t) = \sum_{t=1}^{\infty} tp^T (R^{t-1} - R^t) \vec{1} = \\ &= p^T \left(\sum_{t=1}^{\infty} t \cdot R^{t-1} \right) (I - R) \vec{1} = p^T (I - R)^{-2} (I - R) \vec{1} = p^T (I - R)^{-1} \vec{1} \end{aligned} \quad (\text{A.6.8})$$

Références

- [1] ABRAHAM B. et CHUANG A.
Outlier Detection and Time Series Modeling
Technometrics, 1989, vol. 33, n°2, pp. 241-248
- [2] ALT, F.B.
Multivariate Quality Control: State of the Art
Transactions of the 1982 ASQC Quality Congress, 1982, Detroit, MI, pp.886-893
- [3] APARISI F.
Hotelling's T² Control Chart with Adaptive Sample Size
Int. J. Prod. Res., 1996, vol. 34, no. 10, 2853 - 2862
- [4] BAILLARGEON G.
Maîtrise Statistique des Procédés
Les Editions SMG, 1992
- [5] BARNETT V. et LEWIS T.
Outliers in Statistical Data
John Wiley, New York, 1984
- [6] BLAZEK, L.W., NOVIC, B., et SCOTT, D.M.
Displaying Multivariate Data Using Polyplots
Journal of Quality Technology, 1987, 19, pp. 69-74
- [7] BROOK D. et EVANS D.A.
An approach to the probability distribution of CUSUM Run Length
Biometrika, 59, 539-549, 1972
- [8] BRUCE A. et MARTIN A.G.
Leave-k-Out Diagnostics for Time Series
J. R. Stat. Soc., B, 1989, n°3, pp. 363- 424
- [9] CASTAGLIOLA P.
Evaluation of Non-Normal Process Capability Indices Using Burr's Distributions
Quality Engineering, 8(4), 587-593, 1996
- [10] CHAMP C.W. et RIGDON S.E.
A Comparison of the Markov Chain and the Integral Equation Approaches for
Evaluating the Run Length Distribution of Quality Control Charts
Communications in Statistics, 20(1), 191-204, 1991

- [11] CHEN, K.S et PEARN W.L
An Application of Non-Normal Process Capability Indices
Quality and Reliability Engineering International, vol.13, 355-360, 1997
- [12] CHERNICK M.R., MURTHY V.K., NEALY C.D.
Application of bootstrap and other resampling techniques: evaluation of classifier performance.
Pattern Recognition Letters, 1985, vol. 3, pp. 167-178
- [13] CHOW C. K.
On optimum Recognition Error and Reject Trade-off
IEEE Trans. on Information Theory, 1970, vol. 16, n°1
- [14] CHUA M.K. et MONTGOMERY D.C.
Investigation and Characterisation of a Control Scheme for Multivariate Quality Control.
Quality and Reliability Engineering International, 1992, vol. 87, 37-44
- [15] CHUA M.K.
A Control Scheme for Multivariate Quality Control
Unpublished PhD. Dissertation, Arizona State University, 1990
- [16] COCHRAN, W.G.
The Distribution of the Largest of a Set of Estimated Variance as a Fraction of their Totals
Annals of Eugenics, 1941, vol. 16, pp. 468-473
- [17] CROSIER R.B.
Multivariate Generalisation of Cumulative Sum Quality-Control Schemes
Technometrics, 1988, 30(3), pp. 291-303
- [18] CROWDER S.V. et al.
Process Control and Statistical Inference
Journal of Quality Technology, 1997, Vol. 29, No. 2
- [19] CROWDER, S.V.
A Simple Method for Studying Run Length Distributions of Exponentially Weighted Moving Average Control Charts
Technometrics, 1987,29, pp.401-407
- [20] DEMING, W.E.
Quality, productivity, and competitive position.
MIT, Center for Advanced Engineering Study: Cambridge, Mass., 1982.
- [21] DEVIJVER P.A., KITTLER J.
Pattern Recognition: a statistical approach.
Prentice Hall, 1982

- [22] DIXON W.J.
Analysis of extreme values
Annals of Math. Stat., vol.21, pp.488-506, 1950
- [23] DIXON, W. J.
Processing Data for Outliers
Biometrics, pp.74 - 89
- [24] DODGE H. F., ROMIG H. G.
Sampling Inspection Tables
Second Edition, John Wiley & Sons, NY., 1959
- [25] DOGANAKSOY N., FALTIN F.W. et TUCKER W.T
Identification of Out of Control Quality Characteristics in a Multivariate Manufacturing Environment.
Communications in Statistics - Theory and Methods, 1991, 20(9).
- [26] DRAPER N. R. et JOHN J. A.
Influential Observation and Outliers in Regression
Technometrics, 1981, vol. 23, pp. 21-26
- [27] DUBUISSON B.,
Diagnostic et reconnaissance des formes.
Traité des Nouvelles Technologies, série Diagnostic et Maintenance, Hermès, 1990
- [28] DUBUISSON B.
Decisions with rejet options. In Signal Processing V: Theories and applications
Elsevier Science Publishers, 1990, Vol. 3, pp. 1715 - 1718
- [29] EFRON B.
Estimating the error rate of the prediction rule: Improvements on cross validation.
Journal of the American Statistical Association, 1983, vol.78, pp.316-331
- [30] ECKES G.
Objectif Six Sigma.
Village Mondial, 2001
- [31] FOROUTAN I. et SKLANSKY J.
Feature selection for automatic classification of non-gaussian data.
IEEE Trans. on Systems, Man, Cybernetic, 1987, vol. 17, n°2, pp. 187-197
- [32] FOX A. J.
Outliers in Time Series
J. R. Stat. Soc. B, 1972, vol. 34, pp. 350-363
- [33] FUCHS C. et BENJAMINI Y.
Multivariate Profile Charts for Statistical Process Control.
Technometrics, 1994, vol. 36, No. 2.

- [34] FUKUNAGA K.
Introduction to Statistical Pattern Recognition.
Academic Press, New York, 1990
- [35] FUKUNAGA K. et MANTOCK J. M.
Nonparametric Data Reduction
IEEE Trans. on Pattern Analysis and Machine Intelligence, 1984, vol. 6, n°1, pp. 115-118
- [36] GANANADESIKAN R. et KETTENRING J. R.
Robust Estimates, Residuals and Outlier Detection with Multiresponse Data
Biometrics, vol. 28, pp. 81-124, 1972
- [37] GARVIN, D.A
What does 'product quality' really mean?
The Sloan Management Review, Fall 1984, 25-43.
- [38] GILBRETH, F.
Motion Study,
NY, D. Van Nostrand Co., 1911
- [39] GILBRETH, F.
Primer of Scientific Management
NY, D. Van Nostrand Co., 1912
- [40] GRUBBS, F.E.
On Estimating Precision of Measuring Instruments and Product Variability
Journal of American Statistical Association, 1948, 43, pp. 243-264
- [41] GRUBBS, F.E.
Errors of Measurement, Precision, Accuracy, and the Statistical Comparison of Measuring Instruments
Technometrics, 1973, 15, pp. 53-66
- [42] GRUBBS, F.E.
Sample criteria for testing outlying observations
Annals of Math. Stat., vol.21, pp. 27-58, 1950
- [43] GRUBBS F. E.
Procedures of Detection Outlying Observations in Samples
Technometrics, vol.4, n°1, pp.1-21, 1969
- [44] GUNTER, B.H
The use and Abuse of C_{PK}
Quality Progress, 22(3), pp.108-109

- [45] HARRY M. et SCHROEDER R.
Six Sigma, the Breakthrough Management Strategy. Revolutionizing The
World's Top Corporations
Bantam Doubleday Dell Books, 1999.
- [46] HAWKINS D.M.
Regression Adjustment for Variables in Multivariate Quality Control.
Journal of Quality Technology, 1993, vol. 25, No. 3.
- [47] HAWKINS, D.M.
Diagnostic for Use With Regression Recursive Residuals
Technometrics, 1991, vol.33, no.2, pp. 221-234
- [48] HAWKINS D.M.
Identification of Outliers
Chapman and Hall, London, 1980
- [49] HAWKINS D.M.
The detection of errors in multivariate data using principal components
J. Amer. Stat. Assoc., vol. 69, n°346, pp.340-344, 1974
- [50] HEALY, J.D.
A Note on Multivariate CUSUM Procedures
Technometrics, 1987, 29, pp. 409-412
- [51] HOLMES, D.S. et MERGEN, A.E.
Improving the Performance of the T² Control Chart
Quality Engineering, 1993, 5, pp.619-625
- [52] HOSKULDSSON A.
PLS Regression Methods
J. Chemometrics, 1988, 2, pp. 211-228
- [53] HOTELLING H.H.
Multivariate Quality Control Illustrated by the Air Testing of Sample
Bombsights
Techniques of Statistical Analysis, 1947, pp.111-184
- [54] HOUSHMAND A. A. et JAVAHERY A.
Multivariate Ridge Residual Charts.
Quality Engineering, 1998, 10(4), 617-624.
- [55] JACKSON J. E.
A User's Guide to Principal Components
John Wiley and Sons Inc, New York, 1991
- [56] JACKSON J. E. et MUDHOLKAR J. F.
Control procedure for residuals associated with principal components analysis
Technometrics, 1979, 21 (3), pp. 341-349

- [57] JAIN, A.K., DUBES R.C., CHEN C.C.
Bootstrap techniques for error estimation
IEEE Trans. on Pattern Analysis and Machine Intelligence, 1987, vol.9, n°5,
pp. 628-633
- [58] JAIN, A.K., MOREAU J.V.
Bootstrap technique in cluster analysis.
Pattern Recognition, 1987, vol.20, n°5, pp.547-568
- [59] JAIN A.K. et RAMASWAMI M.D.
Classifier Design for Partzen Windows.
In Pattern Recognition and Artificial Intelligence, Gelsema E.S. and Kanal L.N.
(eds), Elsevier Science, 1988, pp. 211-228
- [60] JAIN A.K. et CHANDRASEKARAN B.
Dimensionality and sample size consideration in pattern recognition practice
Handbook of Statistics, 1982, vol. 2, Krishnaiah P.R. and Kanal L. N. (Eds.),
North-Holland, pp. 547-568
- [61] JOUSSELIN A.
Extension de méthodes de reconnaissance statistique des formes pour
l'applicabilité au diagnostic
Thèse de Docteur-Ingénieur, Université de Technologie de Compiègne, 1985
- [62] KIPARISSIDES C., VERROS G. et MacGREGOR J.F.
Mathematical Modelling, Optimisation and Quality Control of High Pressure
Ethylene Polymerization Reactors
*Journal of Macromolecular Science, Reviews in Macromolecular Chemistry and
Physics*, 1993, C33 (4), pp.437 - 527
- [63] KOBI, A.
Diagnostic de processus continus: Application à la détection des valeurs
aberrantes dans les signaux d'entrée et de sortie de systèmes
Thèse: Spécialité Automatique; Institut National Polytechnique de Lorraine;
1994; N°: URA 821
- [64] KOSANOVICH K. A., DAHL K. S., PIOVOSO, M. J.
Improved Process Understanding using Multiway Principal Component Analysis
Industrial & Engineering Chemistry Research, 1996, 35 (1), pp. 138-146
- [65] KOURTI T. et MacGREGOR J.F.
Multivariate SPC Methods for Process and Product Monitoring.
Journal of Quality Technology, 1996, vol. 28, No. 4.
- [66] KOURTI T., LEE J., MacGregor J. F.
Experiences with industrial applications of projection methods for multivariate
statistical process control
Computers and Chemical Engineering, 1996, 20, pp. 747-752

- [67] KRESTA J., MacGREGOR J. F., MARLIN T. E.
Multivariate Statistical Monitoring of Process Operating Performance
Can. J. Chem. Eng., 1991, 69, pp. 35-47
- [68] LAFAYE DE MICHEAUX D.
Indicateurs multivariés pour une maîtrise globale de processus
Actes du Congrès Qualita, Annecy, France, 2001, pp. 143-151
- [69] LAFAYE DE MICHEAUX D.
Cartes de contrôle multidimensionnelles
Collège Statistique du MFQ, septembre 1998
- [70] LEBART L., MORINEAU A. et PIRON M.
Statistique exploratoire multidimensionnelle
Dunod, Paris, 1997, ISBN 2 10 004001 4
- [71] LEDOLTER J.
The effect of Additive Outliers on the Forecasts from ARIMA Models
Int. J. Forecast., 1989, vol.5, pp. 231 - 240
- [72] LOWRY C.A. et MONTGOMERY D.C.
A Review of Multivariate Control Charts
IIE Transactions, 1995, 27, pp. 800-810
- [73] LOWRY C.A., et al.
A Multivariate Exponentially Weighted Moving Average Control Chart
Technometrics, 1992, 34, pp. 46-53
- [74] LUCAS J.M. et SACCUCCI M.S.
Exponentially Weighted Moving Average Control Schemes: Proprieties and Enhancements
Technometrics, 32, 1-12, 1990
- [75] MacGREGOR J.F.
Using On-Line Process Data to Improve Quality: Challenges for Statisticians
International Statistical Review, 1997, 65, 3, 309-323
- [76] MANDEL, B.J.
The regression Control Chart
Journal of Quality Technology, 1969, 1, pp.1-9
- [77] MARCON M.F. et YOUNG J.C.
Multivariate Statistical Process Control. What, Why and When - the How Comes Later
Quality Progress
- [78] MASON R.L., et al.
Assessment of Multivariate Process Control Techniques
Journal of Quality Technology, April 1997, Vol. 29, No. 2

- [79] MASON R.L., et YOUNG J.C.
Why Multivariate Statistical Process Control?
Quality Progress, 1998
- [80] MASON R. L., TRACY N.D. et YOUNG J.C.
Decomposition of T^2 for Multivariate Control Chart Interpretation.
Journal of Quality Technology, 1995, vol. 27, No. 2.
- [81] MASON R. L., TRACY N.D. et YOUNG J.C.
A Practical Approach for Interpreting Multivariate T^2 Control Chart Signals.
Journal of Quality Technology, 1997, vol. 29, No. 4.
- [82] MASON R. L. et YOUNG J.C.
Improving the Sensitivity of the T^2 Statistic in Multivariate Process Control.
Journal of Quality Technology, 1999, vol. 31, No. 2.
- [83] MASTRANGELO, C.M. et al.
Statistical Process Monitoring with Principal Components
Quality and Reliability Engineering International, 1996, 12, pp.203-210
- [84] McCabe G.P.
Principal variables.
Technometrics, 1984, vol. 26, n°2, pp. 137 - 144
- [85] MONTGOMERY D.C.
Introduction to Statistical Quality Control
Jhon Wiley and Sons, Inc
- [86] MONTGOMERY D.C. et FRIEDMAN D.J.
Prediction Using Regression Models With Multicollinear Predictor Variables
IIE Transactions, 1993, vol. 25, n° 3
- [87] MUIRHEAD C. R.
Distinguishing Outliers Types in Time Series
J. Roy. Stat. Soc., B, 1986, vol. 48, pp. 39 - 47
- [88] NEOGI D. et SCHLAGS C. E.
Multivariate statistical analysis of an emulsion batch process
Industrial & Engineering Chemistry Research, 1998, 37 (10), pp. 3971-3979
- [89] NOMIKOS P. et MacGREGOR J. F.
Multivariate SPC Charts for Monitoring Batch Processes
Technometrics, 1995, 37 (1), pp. 41-59
- [90] PAGE, E.S.
Continuous Inspection Schemes,
Biometrika, 41, 1954

- [91] PEARSON, E.S.
The Application of Statistical Methods to Industrial Standardization and Quality Control
British Standards Institution, London, UK, 1935
- [92] PEARSON, E.S.
Some Notes on W.A. Shewhart's Influence on the Application of Statistical Methods in Great Britain
Industrial Quality Control, 1967, pp. 81-83
- [93] PEREZ W. M.
Six Sigma - Understanding the concepts, implications and challenges
ISBN 85-7303-247-2
- [94] PIGNATIELLO, J.J. et RUNGER, G.C.
Comparisons of Multivariate CUSUM Charts
Journal of Quality Technology, 1990, 22(3), pp. 173-186
- [95] PILLET M.
Appliquer la Maîtrise Statistique des Procédés (MSP/SPC)
Les Editions d'Organisation, 1997
- [96] POLANSKY A.M, CHOU Y-M. et MASON R.L
Estimating Process Capability Indices for a Truncated Distribution
Quality Engineering, 11(2), 257-265, 1998-1999
- [97] PRABHU S.S. et RUNGER G.C.
Designing a Multivariate EWMA Control Chart
Journal of Quality Technology, 1997, 29(1), pp. 8-15
- [98] PRABHU S.S. et RUNGER G.C.
Analysis of a Two-Dimensional Markov Chain
Communications in Statistics - Simula., 1996, 25(1), pp. 75-79
- [99] QUESSENBERRY C.P.
SPC Methods for Quality Improvement
New York, Wiley, 1997
- [100] RAO C. R.
The use and interpretation of principal components analysis in applied research
Sankhya A., vol.26, pp.329 - 358, 1964
- [101] RAO, C.R
Advanced statistical methods in biometric research
Hafner Publishing Company, Darien, Conn, 1970.

- [102] RAUDYS S.J., JAIN A.K.
Small Sample Size Effects in Statistical Pattern Recognition: Recommendation for Practitioners
IEEE Trans. on Pattern Analysis and Machine Intelligence, 1991, vol.13, n°3, pp. 234-252
- [103] ROBERTS, S. W.
Control Chart Tests Based on Geometric Moving Averages.
Technometrics 1, 239-250, 1959
- [104] RUNGER G.C. et MONTGOMERY D.C
Multivariate and Univariate Process Control. Geometry and Shift Directions
Quality and Reliability Engineering International, 1997, vol. 13, 153-158
- [105] RUNGER G.C, ALT F.B. et MONTGOMERY D.C.
Contributors to a Multivariate Statistical Process Control Chart Signal.
Communications in Statistics - Theory and Methods, 1996, 25(10).
- [106] SACCUCCI M.S. et LUCAS J.M.
Average Run Lengths for Exponentially Weighted Moving Average Control Schemes Using the Markov Chain Approach,
Journal of Quality Technology, 22, 154-162, 1990
- [107] SCHOLZ, F.W et TOSCH, T.J.
Small Sample Uni and Multivariate Control Charts for Means
Proceedings of the American Statistical Association, Quality and Productivity Section, 1994
- [108] SHEWHART W. A.
Economic Control of Quality of Manufactured Product
New York, , D. Van Nostrand Co., 1931
- [109] SIMON J. C.
La reconnaissance des formes par algorithmes
Masson, Paris, 1984
- [110] SMOLARZ A.
Un algorithme de discrimination adaptatif avec rejet
R.A.I.R.O. APII, 1987, vol. 21, n°5, pp. 449-474
- [111] SOMERVILLE S.E. et MONGOMERY D.C.
Process Capability Indices and Non-Normal Distributions
Quality Engineering, 9(2), 305-316, 1996-1997
- [112] SOUVAY, P. et MAGNIEN P.
SPC, symmetries et coaxialités
Technologies et formations, 54, pp. 33-36, 1998

- [113] SULLIVAN, J.H., et WOODALL W.H.
A Comparison of Multivariate Control Charts for Individual Observations
Journal of Quality Technology, 1996, 28(4), pp. 398-408
- [114] SULLIVAN, J.H., et WOODALL W.H.
Adapting Control Charts for the Preliminary Analysis of Multivariate Observations
Communications in Statistics, 1998, 27(4), 953-979
- [115] STATGRAPHICS 3.0
Statistical Graphic System, 1988, Rockville, MD: STSC
- [116] TAGUCHI G.
Introduction to Quality Engineering: Designing Quality into Products and Processes
Asian Productivity Organization, available in the USA from American Supplier Institute, 1986, Dearborn, MI, USA.
- [117] TAGUCHI G.
System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Costs
Vols. 1 & 2, UNIPUB/Kraus, International Publications, White Plains, NY, USA.
- [118] TAGUCHI G.
Taguchi on Robust Technology Development: Bringing Quality
Engineering Upstream, ASME Press, 1993, New York NY.
- [119] TAGUCHI G. et CLAUSING D.
Robust Quality
Harvard Business Review, 1990, Jan.-Feb.
- [120] TAGUCHI G. et KONISHI S.
Taguchi Methods, Orthogonal Arrays and Linear Graphs,
American Supplier Institute, Dearborn, MI, USA, 1987
- [121] TAGUCHI G. et YOKOYAMA Y.
Taguchi Methods: Design of Experiments
American Supplier Institute, Dearborn MI, in conjunction with the Japanese Standards Association, Tokyo, Japan, 1994
- [122] TAYLOR, F.W.
Scientific Management.
Harper & Row Ltd.: New York, 1974
- [123] TRACY N.D., YOUNG J.C. et MASON R.L.
Multivariate Control Charts for Individual Observations
Journal of Quality Technology, 1992, 24(2), pp. 88-95

- [124] TRACY N.D., YOUNG J.C. et MASON R. L.
A Bivariate Control Chart for Paired Measurements.
Journal of Quality Technology, 1995, vol. 27, No. 4.
- [125] WEISS S.M.
Small Sample Error Rate Estimation for k-NN Classifiers
IEEE Trans. on Pattern Analysis and Machine Intelligence, 1991, vol. 13, n°3,
pp.258-289
- [126] WESTERHUIS J. A., GURDEN S. P. et SMILDE A. K.
Generalized contribution plots in multivariate statistical process monitoring
Elsevier, Chemometrics and Intelligent Laboratory Systems, 2000, 51, pp. 95-114
- [127] WIERDA S.J.
Multivariate Statistical Process Control - Recent Results and Directions for
Future Research
Statistica Neerlandica, 1994, Vol. 48, nr. 2, pp.147-168
- [128] WIERDA, S.J.
Multivariate Statistical Process Control
*Groningen Theses in Economics, Management and Organisation. Wolters-
Noordhoff, Groningen, The Netherlands, 1994*
- [129] WISE B. M. et RICKER N. L.
Recent Advances in Multivariate Statistical Process Control: Improving
Robustness and Sensitivity
*K. Najim, J. P. Parbary (Ed.), IFAC International Symposium, ADCHEM'91,
Toulouse, France, 1991, pp.125 - 130*
- [130] WOLD S.
Cross-validatory estimation of the number of components in factor and principal
component models
Technometrics, 1978, 20, pp. 397-405
- [131] WOODALL W.H. et MONTGOMERY D.C.
Research Issues and Ideas in Statistical Process Control
Journal of Quality Technology, 1999, Vol. 31, No.4
- [132] WOODALL W.H. et MONTGOMERY D.C.
A Discussion on Statistically-Based Process Monitoring and Control
Journal of Quality Technology, 1997, Vol. 29, No. 2
- [133] YOON S. et MacGREGOR J. F.
Fault diagnosis with multivariate statistical models part I: using steady state fault
signatures
Elsevier, Journal of Process Control, 2001, 11, pp. 387-400
- [134] ZHANG G.X.
A New Type of Control Charts and a Theory of Diagnosis with Control Charts.
World Quality Congress Transactions. American Society for Quality Control

Résumé

Les travaux présentés dans ce mémoire se situent dans le cadre de la maîtrise statistique des processus (MSP). Les principales cartes de contrôle univariées et multivariées sont présentées dans le chapitre I. Un résumé non-exhaustif des méthodes de détection de la cause d'une situation hors contrôle dans un processus multivarié est donné dans le chapitre II. Une nouvelle méthodologie de contrôle et de diagnostic de processus – FNAD (Filtrage Numérique et Analyse Discriminante) – a été proposée dans le chapitre III. Un exemple pratique d'utilisation de la méthode FNAD a été donné pour le diagnostic d'un processus de fabrication du polyéthylène de faible densité. Deux nouvelles cartes de contrôle - la carte de contrôle par filtrage numérique (CCFN) et la carte de contrôle spectrale (CCS) ont été proposées dans le chapitre IV. Le principe de construction de la carte CCFN repose sur l'utilisation des équations récurrentes définissant les filtres numériques RII (Réponse Impulsionnelle Infinie) ou ARMA (AutoRegressive Moving Average). Les limites de contrôle et l'efficacité de la carte CCFN ont été calculées à l'aide de simulations. Le principe de contrôle de la carte de contrôle spectrale (CCS) est basé sur l'analogie temps-fréquence: une fenêtre glissante est déplacée dans le domaine temporel et le contenu spectral de la fenêtre est analysé afin de détecter les changements de la moyenne. La formule analytique de la limite de contrôle a été donnée. Les performances de la carte dans la détection des différents dérèglages ont été calculées à partir des simulations et comparées avec celles de la carte EWMA.

Abstract

The work presented here is in the frame of the statistical process control (SPC). We are presenting a brief description of the Statistical Quality Control (SQC) and some theoretical aspects concerning the multivariate control charts in chapter I. Some representative methods used to identify the variable responsible for an out of control situation in multivariate processes are given in chapter II. A new control and diagnosis methodology - DADF (Discriminant Analysis and Digital Filtering) - is proposed in chapter III. The practical utility of the DADF method is analysed through an example taken from a manufacturing process of the low-density polyethylene. Two new control charts - the Digital Filtering Control Chart (DFCC) and the Spectral Control Chart (SCC) are given in chapter IV. The construction of the digital filtering control chart is based on the use of the recursive equations defining IIR (Infinite Impulse Response) filters. The control limits and the performances of the DFCC were calculated using simulations. The construction of the spectral control chart is based on the time-frequency duality: the spectral content of a moving time window is analysed in order to detect weak changes in the process mean. The Average Run Length (ARL) of the spectral control chart was calculated using simulations. A comparison between the performances of the EWMA (Exponentially Weighted Moving Average) control chart and the spectral control chart was carried out.

Spécialité: Sciences et Technologies Industrielles

Mots-Cles:

MSP, Diagnostic, Filtrage Numérique, Analyse Discriminante, Classification, Carte de Contrôle par Filtrage Numérique, Carte de Contrôle Spectrale.

Key words:

SPC, Diagnosis, Digital Filtering, Discriminant Analysis, Clustering, Control Charts based on Digital Signal Filtering, Spectrum Control Chart.

**ISTIA (l'Institut des Sciences et Techniques de l'Ingénieur d'Angers)
Laboratoire LASQUO (Laboratoire d'Assurance Qualité et Organisations)
62, avenue Notre Dame du Lac, 49000 Angers (France)**